DOCUMENT RESUME

ED 267 098 TM 860 183

AUTHOR Gage, N. L.
TITLE Hard Gains in the Soft Sciences: The Case of Pedagogy.
INSTITUTION Phi Delta Kappa, Bloomington, Ind. Center on Evaluation and Research.
REPORT NO ISBN-O-87367-719-6
PUB DATE 85
NOTE 76p.
PUB TYPE Books (010) -- Viewpoints (120) -- Information Analyses (070)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Adults; *Educational Research; *Meta Analysis; *Outcomes of Education; *Research Methodology; Research Problems; Scientific Methodology; Teacher Education; Teacher Effectiveness; *Teacher Improvement; Teachers; Teaching (Occupation); *Teaching Methods

ABSTRACT

This monograph updates a 1978 publication titled "The Scientific Basis of the Art of Teaching." Based on lectures presented during the 1983-84 academic year, this work discusses research on teaching to provide a substantial scientific basis for the art of teaching and for the education of teachers in that art. For several decades, one research reviewer after another concluded that relationships between teaching practices and educational outcomes were not evident. The negative consensus was largely due to a simple but widespread fallacy: the fallacy of taking seriously the statistical significance of statistically weak single studies. Recent meta-analysis has improved the procedures for drawing conclusions from any set of studies of relationships between roughly similar variables. Divided into three chapters, this book: (1) reviews the research gains made evident through meta-analysis; (2) discusses the application of these gains for the improvement of teaching practices and teacher education; and (3) considers the process-product, ethnographic-sociolinguistic, and other paradigms for research on teaching to improve these gains. Six pages of references conclude the document. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED267098

HARD GAINS IN THE SOFT SCIENCES

The Case of Pedagogy

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

D. Klower

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A CEDR Monograph
Phi Delta Kappa

N. L. Gage

TM 860 183

Recipients of Phi Delta Kappa's
Biennial Award for Meritorious Contributions
to Education Through Evaluation, Development,
and Research

Each recipient of the Meritorious Contribution Award wrote a monograph. The titles of those monographs accompany the authors' names.

1975 — **John I. Goodlad**
The Uses of Alternative Theories of Educational Change

1977 — **John C. Flanagan**
Project Talent and Related Efforts to Improve Secondary Education

1979 — **Roald F. Campbell**
The Development of the Professorship in Educational Administration

1981 — **N. L. Gage**
Hard Gains in the Soft Sciences: The Case of Pedagogy

1983 — **Robert L. Thorndike**
Intelligence as Information Processing: The Mind and the Computer

HARD GAINS
IN THE
SOFT SCIENCES
The Case of Pedagogy

N. L. Gage

PDK Award
for Meritorious Contributions to Education
Through
Evaluation, Development, and Research

A Publication of Phi Delta Kappa's
Center on Evaluation, Development and Research
Bloomington, Indiana

Cover design by Char Dapena

© 1985 Phi Delta Kappa
All Rights Reserved
Library of Congress Catalog Card Number 84-62826
ISBN 0-87367-719-6
Printed in the United States of America

To Maggie,
With Love

Contents

Foreword ... ix
Preface ... xi

I. Reviewing the Gains ... 1

What Kind of Teaching? ... 3
What Is an Art of Teaching? ... 4
What Is a Scientific Basis? ... 7
The Value of Weak Relationships ... 11
A Hypothetical Experiment ... 11
Actual Medical Experiments ... 12
Present-Day Views of the Evidence ... 15
Studies of Intact Patterns of Teaching ... 16
Open Education ... 16
Other Intact Patterns ... 18
Studies of Discrete Teaching Practices ... 19
Experiments with Correlationally-Derived Teaching Practices ... 20

II. Applying the Gains ... 25

The Importance of Pedagogy ... 26
The Possibility of Improving Teaching ... 27
Are the Findings Obvious? ... 28
Evidence on the Improvability of Teaching ... 30
The Hawthorne Effect? ... 34
The Observational-Correlational-Experimental Loop and Teacher Education ... 37

III. Improving the Gains ... 41

What Are Paradigms? ... 41
Paradigms for Research on Teaching ... 44
The Process-Product Paradigm ... 44
Other Paradigms ... 46
The behavior modification paradigm ... 46
The interactive educational technology paradigm ... 46
The instructional design paradigm ... 46
The intact teaching styles paradigm ... 47

The Persistence of Classroom Teaching 48
The Technological-Deliberative Distinction 50
The Complementary Uses of Central Tendency and Variability 51
The Ethnographic-Sociolinguistic Paradigm 53
A Time Perspective 56
References 61

Foreword

Since 1975 Phi Delta Kappa has recognized an outstanding scholar in education through its Biennial Award for Meritorious Contributions to Education through Evaluation, Development, and Research. N. L. Gage was the fourth recipient of this award at Phi Delta Kappa's 38th Biennial Council in Chicago in 1981.

The award is significant for two primary reasons. First, the selection process involved nominations from among PDK's 590 chapters, screening by an awards committee, and final approval by Phi Delta Kappa's Board of Directors. That N. L. Gage's scholarly credentials, his record of accomplishments, and his leadership in educational research stood out above all nominees for the award in 1981 is testimony to the high esteem with which he is regarded by his colleagues in Phi Delta Kappa. Second, the award is given for a pattern of achievements in educational research over time, not for a specific study. Thus in recognizing Gage, PDK is holding his work, indeed his career, up to the fraternity as a standard of excellence.

He is the Margaret Jacks Professor of Education at Stanford University, where he also holds a courtesy professorship in psychology. The *Handbook of Research on Teaching* (1963), which he edited, is a classic in the field of educational research.

A co-founder and co-director of Stanford University's Center for Research and Development in Teaching, Gage has been a prolific writer and editor. He is the author of many articles and several books. He has served 17 professional publications as editor, consulting editor, or editorial board member. He is the founding editor of *Teaching and Teacher Education: An International Journal of Research and Studies* (Pergamon Press). He has held office in over 20 professional organizations, including the presidency of the American Educational Research Association and the presidency of the Educational Psychology Division of the American Psychological Association. Gage has also served in educational research programs in Canada, Australia, Indonesia, England, and West Germany. He has been a fellow of the Center for Advanced Study in the Behavioral Sciences, a Guggenheim Fellow, and a visiting fellow of Brasenose College, Oxford University.

Professor Gage has been an active Kappan, both in his local chapter at Stanford University and in national programs, since 1947. This meritorious award monograph is the most recent of his writings on the contributions of science to the art of teaching.

Willard R. Duckett
Phi Delta Kappa

Preface

In 1978 I published a little book titled *The Scientific Basis of the Art of Teaching*. In its hundred pages I argued that such a scientific basis already exists.

The present work aims to update that argument. Some of what I said in 1978 turned out to be wrong, and I shall attempt to correct my errors. Some of it turned out to be right, and I shall present the new arguments and evidence that strengthen the old. Some of what I said lay in the area between right and wrong; those matters I shall attempt to clarify and sharpen.

This work is based on lectures I gave during the academic year 1983-84. I am grateful to those who arranged for me to give the lectures at their institutions: Dr. Harry Judge, Oxford University; Professor Maurice Galton, University of Leicester; Dean Patricia Albjerg Graham, Harvard University; and Professor Gorän Hermerén, the Conference on "Teaching: Art or Science?" sponsored by the Swedish National Board of Colleges and Universities.

For criticisms of an earlier draft, I am grateful to Clare Burstall, Courtney Cazden, Jeanne Chall, Theodore Coladarci, Barak V. Rosenshine, and Lee S. Shulman.

N. L. Gage
Stanford University

I
Reviewing the Gains

The best result of the discussion which has raged so long about the relative educational value of the main branches of learning is the conviction that there is room for them all in a sound scheme, provided that right methods of teaching be employed. It is not because of the limitation of their faculties that boys of eighteen come to college, having mastered nothing but a few score pages of Latin and Greek, and the bare elements of mathematics. Not nature, but an unintelligent system of instruction from the primary school through the college, is responsible for the fact that many college graduates have so inadequate a conception of what is meant by scientific observation, reasoning, and proof. It is possible for the young to get actual experience of all the principal methods of thought. There is a method of thought in language, and a method in mathematics, and another of natural and physical science, and another of faith. With wise direction, even a child would drink at all these springs. The actual problem to be solved is not what to teach, but how to teach.

—From the Inaugural Address of Charles William Eliot as President of Harvard College, October 19, 1869 (Emphasis added).

Teaching is the central process of education. It comes after economic resources, physical facilities, and curriculum have been developed, and it is intended to lead toward learning. It is the social function through which societies foster achievement of educational objectives. Beyond that, teaching merits study because it uses up a lot of money. Furthermore, teaching can

have great personal and emotional significance: A good teacher can make us happy and comprehending; a bad teacher can make us miserable and confused. And teaching is intriguing as an object of scientific or artistic contemplation in its own right, worth trying to understand for the same reasons that we study the heavens or the impressionists.

But there were several years in the late 1960s and early 1970s when the first James S. Coleman Report in 1966, the reanalyses of it edited by Frederick Mosteller and Daniel Patrick Moynihan in 1972, and the Christopher Jencks book on inequality in 1972 were interpreted as meaning that teaching was unimportant. Those works did not compare teaching with no teaching. Rather, they compared educational outcomes associated with teachers of different kinds and concluded that one teacher seemed to have about the same effect as another. Differences in teachers seemed to make only small if any differences in students. The implication, therefore, was that it did not seem worthwhile to try to improve teachers. I shall not go into that story here, except to say that that work used unpromising independent variables, inappropriate dependent variables, and endlessly arguable correlational methods for determining causal relations between the independent and dependent variables. Research and thinking since then have made most educational and social researchers reject that implication. The Coleman data and similar data were simply inadequate as a basis for any conclusions about the amount of variance in educational outcomes causally associated with differences between teachers on dimensions that might be expected to make a difference. So let us put that episode aside and turn to the questions that I do want to consider.

First, what kind of teaching are we talking about? What do we mean by "teaching" for the present discussion?

Second, what do we mean by "the art of teaching"? As you will note, I am asserting that teaching is an art. But we need to clarify what we mean when we refer to teaching as an art.

Third, what do we mean by "a scientific basis"? Can an art have a scientific basis? If teaching is an art, in what sense can it have a scientific basis?

Fourth, what kind of evidence justifies the proposition that there is already some scientific basis for the art of teaching? At this point we shall, of course, be getting to the heart of the argument. We shall look at examples of what research has yielded, especially during the years since 1978. In looking at the evidence, we shall also need to discuss ways of looking at research evidence in the behavioral sciences, because those ways have changed radically in the last few years.

What Kind of Teaching?

What kind of teaching are we talking about? I have in mind classroom teaching, especially as it goes on in elementary and secondary schools in the United States and in many other countries where teaching has been observed. Classroom teaching does not mean lecturing, exclusively, or discussing, exclusively, or tutoring, exclusively. Rather, it means the combination of all these ways of teaching plus the classroom recitation, or relatively rapid-fire teacher questioning and pupil responding, and so-called seatwork. In addition, classroom teaching includes a variety of managerial activities that keep the whole process moving along in an orderly way.

So I am not dealing with anything very new or highly technological or revolutionary. I am considering what all of us experienced during most of the hours, days, and years we spent in classrooms between our sixth and eighteenth years.

Thus, I am not going into programmed instruction, computer-assisted instruction, Bloom's Mastery Approach, the Keller Plan, Individually Prescribed Instruction, the Program for Learning According to Needs, DISTAR (the Direct Instructional System for Teaching Arithmetic and Reading), and many other approaches to teaching that have been developed in recent decades. What all of these not-to-be-discussed approaches have in common is the requirement that teachers depart radically from classroom teaching of the traditional and customary sort. These radical departures require development of special instructional materials, such as programs, modules, or units. They also require the development of special kinds of test materials suitable for each of the programs, modules, or units — test materials in several forms so that tests can be given repeatedly to students who do not pass them on the first or second attempt. These approaches also require the teacher to adhere strictly to patterns of teaching that have been carefully worked out by the developers of these approaches and that severely curtail the teacher's freedom and flexibility.

Some of these radical approaches have excellent records in improving student achievement and attitudes. Why, then, are they not the focus of our present attention? Because, in my view, they will not prevail in the long run. For reasons that are only beginning to be understood, classroom teaching has turned out to be extremely resistant to the attempts of the last 100 years, if not before, to introduce radical changes in what it calls for. As will be noted below, Cuban (1984) has provided historical documentation

of the robustness of classroom teaching — that is, its apparent impregnability in the face of the efforts of progressive and open educators, on the one hand, and technologists with their programs and modules, on the other. It may be that new attempts to replace classroom teaching will succeed where previous movements have failed. For the next decade or two, however, it is likely that classroom teaching will continue to occupy most of the time of the vast majority of our millions of teachers and students, not only in the United States but throughout the world.

What Is an Art of Teaching?

Many writers, including me, say teaching is an art. What does it mean to speak of an "art" of teaching? Teaching is an instrumental or practical art, not a fine art aimed at creating beauty for its own sake. As an instrumental art, teaching is something that departs from recipes, formulas, or algorithms. It requires improvisation; spontaneity; the handling of hosts of considerations of form, style, pace, rhythm; and appropriateness in ways so complex that even computers must, in principle, fall behind, just as they cannot achieve what a mother does with her five-year-old or what a lover says at any given moment to his or her beloved.

I use the term *art* to refer to any process or procedure whose tremendous complexity, resulting from the large number of relevant variables and the interactions among those variables, makes the process in principle one that cannot be reduced to systematic formulas. Thus an artistic process is not, as I use the term, necessarily a good process: successful, creative, productive of beauty, ingenious, versatile — that is, the province of the virtuoso. Rather, just as bad art is nonetheless art, so bad teaching is nonetheless teaching, and teaching is an artistic process without regard to the quality of the teaching. Artistry is inherent in all teaching except that which is completely mechanical. In my view, even the average teacher using conventional practices must, because of the inevitable complexity of the teaching process, engage in artistic activity.

Highet (1958) called his book *The Art of Teaching* because it seemed to him "very dangerous to apply the aims and methods of science to human beings as individuals, although a statistical principle can often be used to explain their behavior in large groups . . ." (p. vii). I must agree, except that statistical results can help a teacher know the averages or trends around which in-

dividual cases will vary, and such knowledge can aid in understanding the individual. I elaborate below on the values of both averages and departures from averages, or variabilities. In short, let us speak of a *scientific basis for the art of teaching* rather than a science of teaching or even an applied science of teaching.

Another conception of teaching as art also restricts the art of teaching to a special kind: teaching that is "evocative" rather than "didactic" (Axelrod, 1973). Evocative teaching employs inquiry and discovery as its major means. Didactic teaching stresses memorization, repetition, and practice. In my view, however, artistry is involved in both evocative and didactic teaching.

Another writer (Wissot, 1979) said that artistic activity seeks "to remain faithful to the representation of particularized emotional experience," is "idiographic in intent," and is intended "to make statements of utmost specificity about the subjective response to the physical world" (pp. 265, 266). Accordingly, Wissot rejected the possibility of a science of teaching, and I agree with him. But he apparently failed to envisage the possibility of a scientific basis for the art of teaching, as opposed to a science of teaching. He saw incongruity between viewing teaching as a subject fit for scientific explanation and viewing it as a subject fit for artistic revelation, whereas I see teaching as amenable to *both* scientific and artistic scrutiny.

Wissot correctly viewed teaching as receiving and responding to "an enormous cascade of stimulation" (p. 268). From that fact he concluded that teachers have good reason for "limiting their reliance on rational anticipation" and relying instead on "their instinctive feel for emerging events" (p. 268). If this view were correct, it would seem futile to attempt to get teachers to change their practices in the directions that empirical research suggests. Yet, as is demonstrated below, such attempts have succeeded. Repeatedly, it has been possible to change teaching practices — not for all teachers and not for all practices, but for enough teachers and enough practices to make an educationally important difference.

Some writers, such as Tom (1984), prefer the metaphor of teaching as a craft to that of teaching as an instrumental art. But the connotations of "craft" seem too devoid of the complexity inherent in teaching. In any case, Tom never defined clearly what he meant by a craft.

Much of the writing about teaching as an art comes across as not very helpful — in fact, as sonorous vacuities. It is full of references to spontaneity, ineffability, unpredictability, idiosyn-

crasy, ingenuity, the touching of souls, the ability to recognize dynamic patterns, losing oneself in the act while remaining aware, perceiving patterns in motion while monitoring oneself, improvisation within a changing field, and the imaginative leap from theory to practice.

The last-mentioned activity — the "leap" — refers to "theory," which may include the kinds of statistical relationships between variables that scientific research on teaching effectiveness can yield. Accordingly, such a conception of artistry implicitly recognizes the value of a scientific basis for the art of teaching. Such a basis will provide the rules, even if they are only weak generalizations, by which artistry can operate in dealing with the unique classroom full of unique students.

Unfortunately, however, the writings of those who have emphasized artistry in teaching have contributed little thus far to the application of scientifically developed knowledge to the actual practice of teaching. References to educational imagination, the orchestration of classroom dialogue, and attention to pattern and expressive nuance — all these resounding statements seem to be saying something important. But the teacher or teacher educator who seeks help from such writing comes away empty-handed. It is easy to tell a teacher to be an artist, but teachers who want to know what they should do to reach that height find few answers in rhetoric about artistry, intuition, and insight.

In recent years scientific method has been applied to study of the aforementioned "leap" from theory to practice, that is, the decisions and judgments that teachers make before and after they enter the classroom. Teachers' thoughts while planning are hard enough to study; their thoughts while interacting with students are even harder to get at. Nonetheless, efforts are being made (Shavelson and Stern, 1981), and as they succeed, we may have more than empty rhetoric to offer teachers to help them act as artists in applying what is learned from the search for a scientific basis for the art of teaching.

In the meantime, we must recognize that most of what we have as a scientific basis consists of statistical relationships between teaching practices and student achievement of the knowledge, understanding, attitudinal, and conduct objectives of education. Applying that scientific basis in the heat of classroom interaction still relies primarily on artistry. But knowledge of the relevant relationships allows teachers to base their artistry on something more than hunch, feeling, intuition, unaided insight, or raw experience.

What Is a Scientific Basis?

Let us turn now to my third question, What is a scientific basis — especially a scientific basis for the art of teaching? My answer here is deliberately modest, perhaps overmodest. At the least, a scientific basis consists of scientifically developed knowledge about the relationships between variables. In any issue of *Science*, the "Reports" almost always deal with relationships between variables. Even reports of single instances, even case studies, usually contain implicit references to relationships between variables.

Notice that this conception of a scientific basis says nothing about theory, nomological networks, systems of postulates and axioms, or hypothetico-deductive relationships. This emphasis on relationships between variables does not deny the desirability of systematic theory; it merely means that, however desirable, systematic theory is not indispensable to any valid conception of science. The grand theories in the natural sciences and even theories of the middle range in the behavioral sciences may represent goals for us to aspire to, but they do not set minimum requirements as to what we must have before we can lay claim to scientific knowledge.

This modest conception of science is intended merely to avoid claiming too much about where research on teaching stands today. As you will note later, we cannot make claims about relationships between variables without creating classifications, or taxonomies, of our variables. In research on teaching, the sets of relationships that we have do possess some semblance of order, organization, rationale, and meaning. They are not mere random collections of correlation coefficients or effect sizes.

The knowledge obtained through scientific method is typically considered to be nomothetic, that is, more or less general across individuals. It is contrasted with idiographic knowledge, which applies to the understanding of a particular event or individual.

The application of nomothetic knowledge from scientific research to idiographic problems occurs in all technological arts and crafts. Thus Schön (1983), in a volume on how professionals think in action, noted that "If it is true that there is an irreducible element of art in professional practice, it is also true that gifted engineers, teachers, scientists, architects, and managers sometimes display artistry in their day-to-day practice" (p. 28).

The bridge designer does not rely solely on nomothetic science in designing his structure. He uses art in applying nomothetic knowledge to the unique problems imposed by the need for a

bridge of a certain length, to carry specific loads, from a bank of one unique geological character to a bank of another unique geological character, in an area with specific winds, temperatures, and susceptibility to earthquakes (Billington, 1984). The physician uses art in applying nomothetic knowledge to the unique problems raised by a diabetic patient with a unique combination of height and weight, dietary habits, age of onset, blood-sugar level, body type, emotional stability, exercise habits, and job requirements. The teacher similarly uses art in applying nomothetic knowledge to the idiographic problems posed by a specific class consisting of unique students of a given age, grade level, ability level, motivational character, ethnicity, socioeconomic status, and initial knowledge, engaged in the study of a particular subject matter, and so on.

The general character of all of these artistic applications of scientific knowledge is the same. The scientific knowledge is nomothetic — consisting of knowledge of relationships between variables (sometimes strong enough to be considered laws). It holds for the more or less general case under specified conditions with all, or all but a few, other variables held constant. The problem to which the nomothetic knowledge is applied is always idiographic: a particular individual case comprising a unique combination of the levels of many other relevant variables. The engineer, the physician, and the teacher are alike in having to use art in applying nomothetic knowledge to idiographic cases; they differ only in the amount and strength of the nomothetic knowledge available for their use. All must compensate for the inadequacies of their nomothetic knowledge by the use of artistry, craftsmanship, and other hard-to-describe procedures and processes.

The nomothetic knowledge becomes more trustworthy as it is confirmed through replications, that is, repetitions of more or less similar investigations. Such confirmations add more to the persuasiveness of the evidence than does the statistical significance, no matter how strong, of a single result. And, as will be shown below, research on teaching has accumulated sizable bodies of confirmed, replicated findings.

At this point the argument is often raised that the relationships between variables in the social and behavioral sciences, including research on teaching, are beset with interaction effects. Such effects make any relationship between two variables change, even in sign, or direction, as the value of a third variable changes. So the relationship between student need for autonomy and student achievement may have one value for students taught by the lecture method and another value for students taught by

the discussion method. Or higher scores on a reading test may be associated with teaching by the linguistic method for students higher in reading readiness and with the whole-word method for students lower in reading readiness. What can make matters even more complicated is the possibility that such three-variable interactions may change as the values of a fourth variable change. And a four-variable interaction may change as a fifth variable changes.

But knowledge of such interactions merely improves knowledge based on dependable main effects, that is, relationships between two variables at a time. The behavioral sciences are full of clusters of studies showing dependable main effects — dozens and even hundreds of studies of the same relationships that have, with high consistency, found the same kind of relationship between pairs of variables. Thus, in the meta-analysis by White (1982) of the relationship between socioeconomic status and achievement, all but a tiny fraction (estimated from a frequency polygon) of 489 correlation coefficients using the student as the unit of analysis were positive (and the mean r was .25), and *all* of the 93 coefficients based on aggregated units of analysis, such as classroom means, were positive (the mean r was .68). Similarly, scholastic aptitude test scores *always* correlate positively with subsequent grade-point averages. Measures of cognitive ability, or mental ability tests, *always* correlate positively with one another. The effects of many kinds of behavior therapy have been confirmed over many scores of replications (Smith and Glass, 1977). The main-effect results of similar replications of studies of teaching will be described below.

The widely cited point made by Cronbach (1975) concerning the importance of interactions has often been misinterpreted. Cronbach showed that, in 17 studies drawn at random from the *Journal of Personality and Social Psychology*, the interaction effects were in the same range of magnitude as the *second*-largest main effect. The largest main effect was considerably larger than the typical interaction effect. Although the largest main effect was typically not the principal concern in these studies (it resulted from variables such as age or practice versus no practice about whose effects there was no prior uncertainty), that fact is beside the point in the context of considering the importance of main effects. Certainly, after being found repeatedly, some main effects become obvious and are no longer the main concern of investigators. That these main effects are well established and expected does not make them weak or unimportant for scientific or practical purposes.

Some writers object to the possibility of a scientific basis for the art of teaching — indeed, of any art based on the behavioral and social sciences, such as counseling, social work, psychotherapy, or the rehabilitation of criminals — on the ground that the phenomena of the social sciences are not natural but man-made, purposeful, and hence unstable. Philosophers (e.g., Gewirth, 1954) have argued convincingly that the social sciences differ radically from the natural sciences in that the laws of the social sciences can be changed by man, while those of the natural sciences cannot be changed; they can only be discovered and perhaps applied by man. Thus, because educational phenomena are man-made, the variables in education will change over the years, and research findings will never prove to be even relatively permanent.

But the changeability of the man-made phenomena is not an all-or-none matter. Changes in the prevalence of various educational practices, such as open education or team teaching or mainstreaming, affect primarily the mean value of the variables involved. They do not necessarily affect the correlations of those variables with other variables, such as outcomes. And it is the study of relationships to which scientific research is largely devoted.

Furthermore, not all of the mean values or relationships are equally changeable. Some may indeed be changed easily, as soon as we become aware of them. Others are so enmeshed in a complex web of strong habits that they are changed only with great difficulty. Thus, when we find that teachers tend to stand closer to pupils they like and that similar relationships are found for tone of voice, eye-gaze, facial expression, speech rate, gestures, and wait-time after a teacher's question or a student's response (Woolfolk and Brooks, 1983), it becomes unlikely that teachers can easily control all the ways in which their nonverbal behavior reflects their attitudes. There are just too many such behaviors for the teacher to keep in mind, monitor, and manipulate. And even if the teachers can control their own nonverbal behavior, they cannot control the kinds of inferences that children draw from such behavior and hence the relationships of their behavior to their effects on children.

In short, the changeability of phenomena or relationships between variables in human affairs is an empirical question and a matter of degree. The fact that such change is possible does not mean that it is highly probable, inevitable, or universal. Any philosophical proof that man can change laws of social science does not mean that men necessarily will change all the laws of

social science. It seems plausible that many of the phenomena of teaching have remained stable over many decades (see Cuban, 1984, discussed below) and that the relationships of those phenomena to others have similarly remained unchanged. But studied efforts to make some changes in those phenomena are likely to have predictable and controllable effects on other phenomena, such as what students learn and feel.

The Value of Weak Relationships

I do not claim that the relationships between variables which have already been identified and confirmed fairly well through research on teaching are strong ones. In correlational terms, the relationships range between .2 and .5.

Are relationships of that magnitude useful? Until recently, many behavioral scientists were likely to look with scorn at values as low as .2. Readers of the Coleman Report and the reanalyses edited by Mosteller and Moynihan considered relationships of that kind to betoken a kind of futility. Teacher characteristics that correlated at that level with student achievement were useless, they thought. When a correlation coefficient of .2 is squared to obtain the percentage of variance in one variable explained by the other, it is seen to explain a mere 4% of the variance.

But, as Rosenthal and Rubin (1979) have shown, "interpretations like this may be very misleading. An appropriate measure of size of effect must be sensitive to the context of the data" (p. 395). Rather than use the Rosenthal-Rubin example, with its correlation of .4, let me use an example with a correlation of only .2.

A Hypothetical Experiment

Table 1 shows the results of a hypothetical experiment in which half of the persons were randomly assigned to a new treatment while the other half were assigned to an old, or standard, treatment. The dependent variable is "good" or "bad." For example, one year after treatment, the subject may be alive or dead, or "still in school" as against "dropped out." Forty percent of the persons who received the old treatment showed "good" results, and 60% of the persons who received the new treatment showed "good" results. The coefficient of correlation for the relationship between treatment and outcome in this table is only .2, and only 4% of variance in the outcome is accounted for by the treatment. Yet the new treatment reduced the percentage of bad outcomes

from 60% to 40%; that is, it caused a decrease of 20% in bad outcomes. Suppose we compare that decrease of 20% with the old treatment's bad-outcome rate of 60%. Then we realize that the new treatment has decreased the rate of bad outcomes by 33% (20/60). And all this is true with a correlation coefficient of only .2.

Table 1

Results of a Hypothetical Experiment

Outcome	*Receiving New Treatment*	*Receiving Old Treatment*
Good	60%	40%
Bad	40%	60%
Total	100%	100%

Correlation (Φ) = .20
Percent of Variance Explained = 4%

Actual Medical Experiments

But these are only hypothetical data. What about the results of actual experiments with an important dependent variable, such as life or death? In 1982 we received the results of a large-scale experiment with a drug called propranolol, which was intended to increase the rate of survival of men who had already had one heart attack. The experiment was conducted with random assignment of about 3,800 such men to either the drug or a placebo. The investigators used double-blind procedures, and presumably the experiment, which cost about $20 million, was conducted impeccably in all other respects. The results of this experiment are shown in Table 2.

After 30 months, 9.5% of the men who had received the placebo had died, while only 7% of those who had received propranolol had died. The drug reduced the percentage of fatalities

by only 2.5%. This kind of result would almost certainly be regarded as trivial in research on teaching. But in the medical arena, the result was regarded as so important that the experiment was discontinued on ethical grounds. The results led to the recommendation that propranolol be used, with a potential saving of about 21,000 lives per year in the United States alone.

Table 2

Results of Beta-Blocker Trial

After 30 Months	*Propranolol (N=1900)*	*Placebo (N=1900)*
Dead	7.0%	9.5%
Alive	93.0%	90.5%
Total	100.0%	100.0%

Correlation (Φ) = .045
Percent of Variance Explained = .2%

Source: Journal of the American Medical Association, 1982, 247, 1707-13.

In 1984 we received news of a similar experiment. This 7- to 10-year-long experiment cost $150 million and dealt with cholesterol lowering through diet and the use of a drug. The results are shown in Table 3. Of the men who got the placebo treatment, 9.8% had had a definite (fatal or nonfatal) heart attack; of the men who got the experimental treatment, only 8.1% had had a definite heart attack. The result was front-page news in the *New York Times* and the *Boston Globe* and led to a cover story in *Time* (March 26, 1984). An article in *Science* (Kolata, 1984) held that the results would "affect profoundly the practice of medicine in this country" (p. 380). But the treatment produced only a 1.7% difference and accounted for only .1% of the variance in heart attacks.

Table 3

Effects of Cholesterol Lowering

Result After Nine Years	*Experimental Treatment* (N=1906)	*Placebo* (N=1900)
Definite Fatal or Nonfatal Attack	8.1%	9.8%
No Definite Fatal or Nonfatal Heart Attack	91.9%	90.2%
Total	100.0%	100.0%

Correlation (Φ) = .03
Percent of Variance Explained = .1%

Source: Journal of the American Medical Association, 1984, 261, 351-64.

Why this excursion into hypothetical and medical research? I am trying to defend the proposition that correlations or differences do not need to be large in order to be important. In education we may not be influencing life or death. But we may be influencing dropout rates, literacy, placement in special classes, love of learning, self-esteem, or a holistic ability to integrate many facts and concepts in a complex way. And independent variables in research on teaching may be no more expensive per student than the drugs used in the medical experiments that are taken so seriously. The implications for practice depend not on the size of effects but on the costs and benefits of any change in practice. Further, as I have noted elsewhere (Gage, 1983), both costs and benefits of changes in teaching should not be measured only in terms of money. Thus effectiveness is not the sole criterion of what a teacher should do. Cost is the other major factor — cost in effort, time, or money, all of which are related (see Levin, 1983). Effective methods that cost too much will tend to become disused. Such disuse has been the fate of many promising

teaching methods and practices. Their effectiveness in improving student achievement and attitudes did not compensate enough for their cost to teachers in effort, time, or money.

Present-Day Views of the Evidence

Thus what I see as a scientific basis for the art of teaching is one or more relationships between things that teachers do and things that students learn. Now, given my definition of teaching, of an art of teaching, and of a scientific basis, can we say that we have a scientific basis for the art of teaching?

Until about 1976, the consensus among most reviewers of the evidence was negative. For several decades, one reviewer after another concluded that relationships between teaching practices and educational outcomes were not evident.

That negative consensus was largely due to a simple but widespread fallacy: the fallacy of taking seriously the statistical significance of statistically weak single studies. A statistically weak study is one whose sample size is small in relation to the magnitude of the expected effect or relationship. The weakness means that the study is unlikely to disprove even a false null hypothesis, given the expected effect-size and the number of cases, such as classrooms. Such a low likelihood means that the weak study is highly likely to lead to the error of concluding that there is no relationship or difference when, in fact, there is one. The statisticians Hedges and Olkin (1980) have shown that the procedure of counting "the number of times the treatment group mean exceeds the control group mean by an amount that is statistically significant" has "extremely low power for the combination of treatment-effect sizes and sample sizes usually found in social science research" (p. 359).

If that kind of vote-counting leads to false conclusions, how should we interpret the results of successive roughly similar studies of the same relationship? In 1976 I used well-known tests of the significance of combined results (Gage, 1976; 1978, pp. 29-30). At the same time, Glass (1976) introduced what he dubbed meta-analysis — a method for the quantitative synthesis of research findings. Glass's meta-analysis has been widely adopted. It has been applied to the synthesis and integration in quantitative form of many clusters of replications of studies of the relationships between variables in the behavioral sciences.

Some dozens of meta-analyses have now been performed in research on teaching. Some behavioral scientists (e.g., Eysenck, in

press) still object strenuously to the whole idea of meta-analysis, and it is clear that the method requires excellent judgment if it is not to yield spurious findings. Yet it seems fair to say that meta-analysis has greatly improved the procedures for drawing conclusions from any set of studies of relationships between roughly similar variables. If so, what do we find when we look at the findings of research on teaching — at the relationships between what teachers do and what students learn?

Studies of Intact Patterns of Teaching

The studies here can be divided into two kinds — those that deal with relatively intact patterns of teaching, or teaching styles, and those that deal with hosts of fairly discrete and relatively independent teaching practices. Let us look first at the studies of intact patterns.

Open Education

One intact pattern of teaching is called "open education," or the "open classroom." That pattern has a long history. It is possible to find its precursors in Socrates, Rousseau, Dewey, and Neill. In the 1960s and 1970s it had a resurgence in England and the United States, and an enormous polemic literature appeared. Then studies comparing what students learned under open and conventional teaching began to appear. In 1977 I reviewed a small and regrettably haphazard set of some seven of those studies and concluded that they showed that students learned less — in reading, mathematics, and writing — under open teaching (Gage, 1978).

A year or so later I learned of a much more thorough review of such studies by Horwitz (1979), who dealt with more than 150 comparisons of the outcomes of open and traditional classrooms. But Horwitz's review made the error mentioned above: He took seriously the statistical significance of weak single studies. And when he counted the votes he found a plurality of the studies falling into the no-difference category.

Peterson (1979) recognized Horwitz's error and performed a meta-analysis. But she made what seemed to be a serious omission: she left out the unpublished dissertations that composed more than half of Horwitz's bibliography.

Therefore, with the support of the Spencer Foundation, we organized a project at Stanford that would look carefully at all

150-plus studies and use improved statistical methods developed by Larry Hedges, who became the project's leader. The results of the meta-analysis by Giaconia and Hedges (1982) illustrate the nature of the evidence on relationships between what teachers do and what students learn. In brief, the results indicated that open education made almost no difference in achievement in reading, mathematics, and language arts. Indeed, if anything, open education was inferior, but only very slightly so, on the average, in bringing about such achievement. The average effect-sizes were only slightly different from zero, and only small majorities of the comparisons favored non-open education.

But open education did better in producing creativity, independence, and favorable attitudes toward school and the teacher. Here 71% of the 23 comparisons on *creativity* favored open education, 68% of the 26 comparisons on *independence* favored open education, and 67% of the comparisons on *attitude toward the teacher* favored open education.

In short, this careful examination of a large, well-nigh exhaustive collection of studies indicated that open education had no advantage or disadvantage when it came to academic achievement but a fairly consistent and moderately substantial advantage when it came to such outcomes as creativity, independence, and attitudes toward the school and the teacher.

Giaconia and Hedges went on to determine what features of the complex known as open education made a difference. To do this, they divided the open education studies into those that yielded large effects and those that yielded small effects on each of the several outcomes on which open education had been evaluated. Then they compared the large-effect and the small-effect studies as to which of the features of open education were associated with the size of the effect. The seven features of open education on which they compared the large-effect and small-effect studies were:

—the role of the child in learning,
—the use of diagnostic evaluation,
—the use of materials to manipulate,
—individualized instruction,
—multi-age grouping,
—open space, and
—team teaching.

When Giaconia and Hedges looked at these specific features of open education, their conclusions were as follows:

> For the nonachievement outcomes, the first four features (role of the child, diagnostic evaluation, materials to manipulate, and indi-

> vidualized instruction) tended to differentiate the studies that yielded larger effects from the studies that yielded smaller effects. . . . For example, *all* of the more effective programs on the outcome of self-concept included these four features, and the vast majority of the larger effect studies of the outcomes creativity and favorable attitude toward school included these four features. On the other hand, only about half of the studies yielding smaller effects on nonachievement outcomes included [these] features (pp. 597-98).

Giaconia and Hedges also concluded that the achievement outcomes were not associated with the various features of open education. Thus, "with the exception of the materials to manipulate feature, the more effective programs were no more likely to include these features than were the less effective programs" (p. 598). And, in general, the last three in the foregoing list of features of open education — namely, open space, multi-age grouping, and team teaching — seemed to be unassociated with the effectiveness of open education in either the nonachievement or the achievement outcomes. They seemed to make no difference at all. The upshot is that open education tends to enhance some important nonachievement outcomes without appreciable sacrifice of achievement outcomes.

Other Intact Patterns

There are other intact patterns of teaching, and the results of research on some of them have been highly consistent. Bloom's Mastery Approach has yielded better achievement test scores in 75% to 80% of the studies in which it has been compared with conventional teaching (Block and Burns, 1976). Direct Instruction of the kind developed by Becker and Engelmann has yielded better achievement than conventional teaching with high consistency (Kennedy, 1978). At the college level, the Keller Plan, or Personalized System of Instruction, has resulted in higher achievement than that produced by conventional courses — again with high consistency (Kulik, Kulik, and Cohen, 1979).

The point is that consistent and substantial differences in achievement and attitude have been found through research on teaching, differences that are a function of differences in intact patterns of teaching. Whether these consistent findings mean that we should hasten to adopt the methods that yield higher achievement or better attitudes is, however, another matter. Here we become involved in what might be called artistic, clinical, or in any case, complex judgments about costs and benefits. Some writers argue that Bloom's Mastery Approach short-

changes the faster students by investing disproportionate resources in helping the slower ones. Some writers argue that the Keller Plan is unsuitable for courses in the humanities, where students have a need and a right to formulate their own importances without having "too arrogant standardization" (Jaynes, 1975, p. 631) imposed on their minds. Some writers argue that the Becker-Engelmann Direct Instruction method imposes on teachers more regimentation than any self-respecting teacher should accept. In any case, in making judgments of this kind, the teacher or the educator must and should go beyond the scientific basis of the art of teaching.

Studies of Discrete Teaching Practices

A second major category of studies of teaching in relation to outcomes deals with large numbers of discrete teaching practices. These practices are typically observed with some kind of systematic observation schedule. The observer determines the frequency with which the teacher does many kinds of things, including structuring, soliciting, reacting, and managing.

Thus various aspects of teacher *structuring* can be observed. Here we determine how often the teacher makes transitions explicit, reviews old lessons, explains the purpose of new ones, outlines ideas, and gives clues as to the importance of ideas.

The observer can also record ways of soliciting, or *asking questions*. Here we determine how many questions are asked, how difficult or complex the questions are, how long the teacher waits for an answer, how the teacher selects a pupil to call upon, and how often the teacher probes, or continues asking questions of the same pupil.

The observer can record how the teacher *reacts* to students' responses. Here we learn how often the teacher gives positive reactions or negative reactions, how long the teacher waits after a student has spoken to allow the student to say more, or how much cognitive information the teacher feeds back to the student to correct the student's errors.

Finally, the observer may also take account of how the teacher *organizes and manages* the class. Here the observer records whether the teacher groups the children and, if so, in what kinds of groups; whether the teacher stays in his or her seat while the pupils are doing seatwork or walks around the room and monitors the seatwork; and how well the teacher engenders "academic

learning time," that is, keeps the pupils engaged in relevant academic tasks at not too high a level of difficulty.

These are just a few of the myriads of variables that have been included in studies of discrete teaching practices. The investigator makes no initial assumption about whether the teacher is using any particular pattern of teaching, such as open or direct instruction.

Then these practices are correlated one at a time with measures of the average achievement and attitude of the classes observed. From the correlation coefficients of the various studies, the investigators and research reviewers then try to infer what kind of teaching practices are associated with higher achievement or more desirable student attitudes.

The literature on this kind of research is voluminous. It has been reviewed and synthesized repeatedly. Perhaps the most recent thorough review is that by Brophy and Good (in press) for the third edition of the *Handbook of Research on Teaching*. Rather than attempt to summarize that review, let me quote their statement that

> . . . we now know much more about teacher effects on achievement than we did in 1963 or even 1973 . . . [and] the fund of available information on producing student achievement (especially the literature relating to the general area of classroom management and to the subject areas of elementary reading and mathematics instruction) has progressed from a collection of disappointing and inconsistent findings to a small but well established knowledge base . . . (ms. p. 132).

What do the results of such a study of teaching practices look like? Figure 1 shows the scatterplot ($r = .40$) obtained by Gage and Coladarci (1980) in a study of 28 teachers of grades 4 through 6 in schools in a predominantly black section of a large city. Note the tendency for residual mean class achievement to be higher in the classes taught by teachers who behaved with higher "conformity to recommendations." These were recommendations that had been derived (Crawford and Gage, 1977) from several previous observational and correlational studies of teaching in the early primary grades.

Experiments with Correlationally Derived Teaching Practices

Now, as everyone knows, correlation does not mean causation. So experiments have been done to determine whether actually

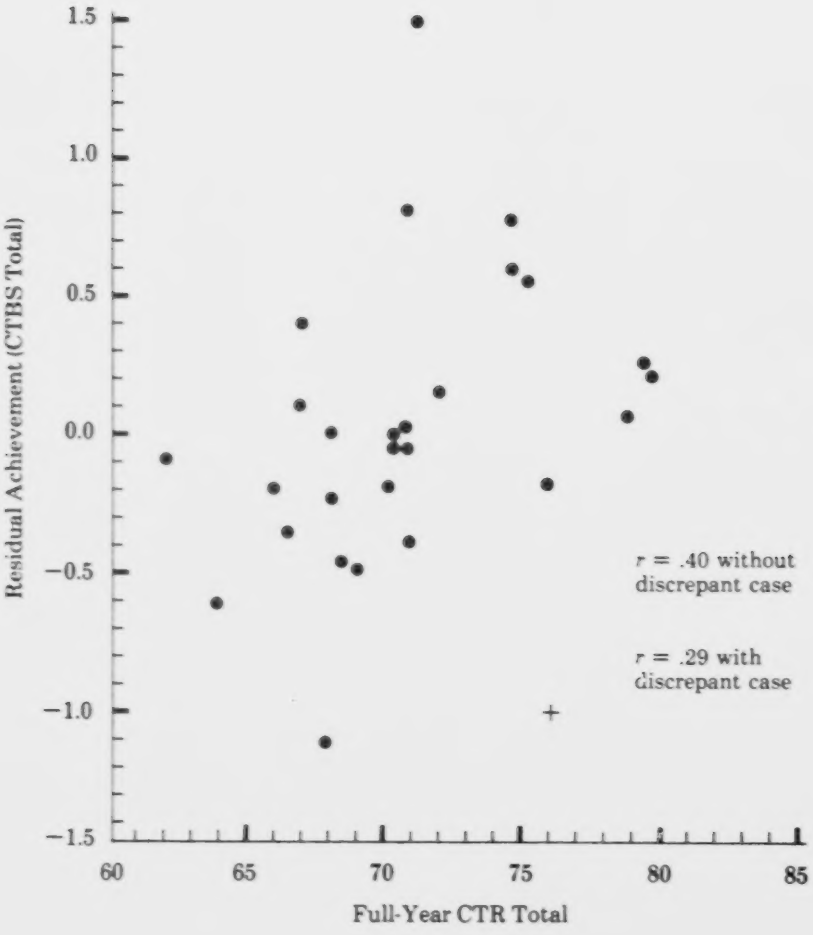

Figure 1. Scatterplot of Full-Year CTR (conformity to recommendations) and Residual Achievement. (Discrepant case is denoted by "+". This teacher was reported by observers to have turned over much of the teaching to an aide.)

Source: Gage and Coladarci (1980).

changing teaching practices in specified ways causes improvement in student achievement, attitudes, and behavior.

These experiments are fairly similar in some novel and important respects. First, the teaching practices that the experimenters tried to change in a specified direction had been derived primarily from previous correlational studies of teaching practices and student achievement. That is, the new teaching was not derived from some ideology, a theory based on laboratory experiments, or some kind of philosophical reasoning. It was derived from correlations between specific teaching practices and student achievement or attitude or both.

Second, these experiments were accompanied by observation in the classrooms. This observation determined the degree to which the desired teaching practices were actually used by the teachers in the experimental group, who got training in using those practices, as compared with the control group, who did not get the training until the experiment was over.

Third, the teachers involved in these experiments were volunteers who agreed to accept random assignment to either the trained group or the control group. The teachers or their schools were then randomly assigned to one group or the other. Such random assignment greatly enhances the validity of the experiments.

Fourth, all teachers were regular teachers who were involved in the experiment for a whole school year or a whole semester, not merely for an hour, a day, or a week or two.

Fifth, the students of the teachers took pretests of achievement and attitude at the beginning of the school year or term and then took comparable posttests at the end.

Sixth, because of the way the teaching practices used in these experiments were derived, they did not represent any revolutionary, arcane, or exotic ways of teaching. They were instead ways of teaching that other teachers in the same kinds of schools had already been seen to be using, teachers in whose classrooms previous investigators had found better-than-average achievement, attitudes, or behavior.

The use of volunteers in these experiments and many others often arouses doubts as to the meaning of the results for teachers in general. Volunteers may have higher motivation to improve or greater interest in research (Rosenthal and Rosnow, 1975). So it is argued that what is learned from volunteers may not apply to the teachers who do not volunteer, and nonvolunteers are often the majority. Two points can be made here. First, research workers have no alternative but to use volunteers, at least in the United States, where teachers cannot be coerced to participate in

educational research. So it becomes a matter of doing research with volunteers or not doing research at all. Second, as results with volunteers become more promising, more and more teachers will volunteer. This phenomenon can be seen in the research on cigarette smoking. In the early days, smokers had to be paid to volunteer for experiments designed to help them quit smoking. As the experiments succeeded, and as the evidence for the desirability of quitting cigarettes became widely known, the proportions of volunteers for such efforts increased tremendously. Nowadays, cigarette smokers pay handsome fees for treatment that at one time they would not have accepted even though it was free. This effect of research success on volunteering, with a corresponding increase in the generalizability of results to previous nonvolunteers, is likely to occur as experiments with teaching practices accumulate a record of improving teacher effectiveness. Whether the results will then still hold up remains to be seen.

Extending a previous review of four such experiments (Gage and Giaconia, 1981), I now can point out that nine experiments of this kind have been reported (Crawford, Gage, Corno, Stayrook, and Mitman, 1978; Anderson, Evertson, and Brophy, 1979; Good and Grouws, 1979; Stallings, Needels, and Stayrook, 1979; Emmer, Sanford, Evertson, Clements, and Martin, 1981; Good and Grouws, 1981; Borg and Ascione, 1982; Emmer, Sanford, Clements, and Martin, 1982; Coladarci and Gage, 1984). In all but one of them, the one by Theodore Coladarci and me, the results were positive and substantial. For example, in the experiment by Crawford and others (1978), 33 third-grade teachers in two school districts were randomly assigned, within each district, to a control group or a training group. The latter group got training over a period of one month. Both groups were observed before, during, and after the training. The results of the observations showed that the training was effective in bringing about significant change on the part of the trained teachers — change in the direction sought through the training. The results of the achievement tests and attitude inventories showed that the classes of the trained teachers had higher mean scores. Indeed, the difference in achievement between the two sets of classes yielded an effect size of .69, which means that the average class of the trained teachers scored at about the seventy-fifth percentile rank of the classes of the untrained teachers.

These experiments suggest that the progression from observational and descriptive studies through correlational studies to experiments has begun to pay off. We are beginning to have evidence that the correlations betoken causal relationships, so that

changing teaching practices *causes* desirable changes in student achievement and attitude. And the changes in achievement are substantial, not trivial. The changes are brought about not by revolutions in teaching practice or school organization but by relatively straightforward and rational attempts to get more teachers to do what more effective teachers have already been observed to be doing.

This kind of research is relatively independent of values. So far, it has been conducted primarily to achieve the values implicit in achievement tests, attitude inventories, and classroom order. But there is nothing in the general approach to prevent the same gains from being made in achieving the values of those who believe in open education or any other kind of education. What is called for is an effort to achieve a scientific basis for the art of teaching.

II
Applying the Gains

Applying what we learn from research on teaching effectiveness is a problem in teacher education. In one sense this problem is the same as that of teaching. But now we are dealing with the teaching of teachers — the teaching that teachers *get* rather than the teaching that teachers *do*. The similarity can be seen when we refer to the kinds of variables entailed.

In Chapter I we dealt with teaching practices as independent variables and student achievement and attitudes as dependent variables. Now we shall deal with teaching practices, broadly defined, as *dependent* variables and with teacher education practices as the independent variables.

Let us take up three questions in turn. First, is teacher education important? In the 1980s the U.S. is going through one of its perennial surges of concern with education, with many study commissions engaged in efforts to diagnose and prescribe for what is wrong with our educational system. Where does teacher education fit in? Is it peripheral or central?

Second, is there evidence that teaching can be taught and learned? Or are teachers born rather than made? Many of us know good teachers who never went through a teacher education program. Does that mean we should leave teachers alone, let them learn on the job, and hope for the best? Or is teaching something that *can* be taught and learned with only a reasonable amount of effort and skill? These questions apply not only to new teachers but to those already on the job.

Third, what kind of research on teacher training is likely to pay off best in view of the peculiar nature of the teacher's job? If teaching is in large degree an art, it cannot be taught in the same way that we train assembly line workers, aircraft mechanics, or

medical technicians. Teachers need more freedom to use their judgment than workers of that kind. But teaching is an art with obligations to the society, with moral imperatives relating to the welfare of students and the body politic. So teachers cannot have the complete autonomy of the creative artist who can choose to be a surrealist, an atonalist, or a dadaist. Thus we need approaches to teacher education that will walk the path between unacceptable regimentation and unacceptable anarchy.

The Importance of Pedagogy

Is teacher education important? Many factors enter into present-day education: finance, buildings, curriculum, and administration. It is hard to know where to begin in trying to improve our educational system. One student of the scene (Howe, 1982) considered several alternative emphases: more equitable funding, desegregation to achieve greater fairness, and vast changes like the voucher system. But he decided that these were "unlikely to produce the improved learning we want within a reasonable period of time" (p. 13). And he then concluded that "the improvement of pedagogy is the most likely source of a real solution to the shortcomings of America's schools" (p. 14).

A similar conclusion appeared in a searching and critical analysis (Stedman and Smith, 1983) of four recent reform proposals for American education: those of the National Commission on Excellence in Education, appointed by Department of Education Secretary Terrel Bell; the Task Force on Education for Economic Growth, established by the Education Commission of the States; the College Entrance Examination Board's Task Force on Academic Preparation for College; and the Twentieth Century Fund's Task Force on Federal Elementary and Secondary Education Policy. The reviewers found a serious omission in all four reports: the reports had "ignored the problem of pedagogy, namely, the question 'How is it taught?' " (p. 95).

Howe's statement and the Stedman-Smith review emphasize the importance of pedagogy. Contrast these emphases with the omission of concern about pedagogy from the four commission reports. The contrast suggests that it is not gratuitous to urge concern with teaching practices and the application of what we have learned through research to the improvement of teaching practices. For whatever reasons, it is evident from the four commission reports that much thinking about education has focused on almost every conceivable part of the enterprise except its cen-

tral process: teaching. So it still seems necessary to support attention to the improvement of teaching as part of the task of improving American education.

The Possibility of Improving Teaching

Can teaching be taught and learned? Can we influence how teachers teach? Can we change the thoughts and actions of teachers? If so, can we consider the change an improvement?

These questions are worth asking because, as everyone knows, teacher education has long been held in low repute. In 1963 Americans were surprised that James Bryant Conant, the former president of Harvard University, had come back from his diplomatic assignment in West Germany to devote himself not merely to education, which was remarkable enough, but to teacher education, a pariah among pariahs. But Conant's effort, *The Education of American Teachers* (1963), had no impact comparable to that of the 1910 Flexner Report on medical education. So, two decades later, universities and political leaders still have scant respect for teacher education.

Even if we give pedagogy a decent respect, we find widespread doubt about the efficacy of teacher education, even — or perhaps especially — among the teachers themselves, who are the graduates of teacher education programs. We hear that the courses teachers take are irrelevant or too theoretical. The courses are full of material about learners and learning, when the prospective teacher wants to know about teachers and teaching. The courses tell about the history, philosophy, sociology, and psychology of education, when the student wants answers to the question, How should I teach? The courses give the prospective teacher more knowledge of the subject matter to be taught, when the teacher may already know far more about that subject than he or she will ever need in teaching third-graders or even twelfth-graders. The evidence that knowledge of the subject is not enough to make a successful teacher is plain to anyone who has ever seen a Ph.D. candidate in mathematics thoroughly confuse a freshman calculus class (Hativa, 1981) or the holder of a bachelor's degree obfuscate the past tense in teaching third-grade reading (Hagberg, 1983).

Thus generations of teacher education students have been given inadequate grounding in how to teach. They have not been taught how to organize a course, how to plan a lesson, how to

manage a class, how to give an explanation, how to arouse interest and motivation, how to ask the right kinds of questions, how to react to students' responses, how to give helpful correction and feedback, how to avoid unfair biases in interacting with students — in short, how to teach.

One reason for the relative neglect of these matters in teacher education programs has been the relative dearth until recently of research-based knowledge about them. Only in the late 1960s and the 1970s did the findings of studies begin to converge. Now those findings have accumulated into bodies of fairly well confirmed knowledge, confirmed in the sense that teaching practices of certain kinds had been found to be associated with higher student achievement and more desirable student attitudes. That there is such knowledge was the argument of my first chapter, and the reader should examine the references cited there for evidence on the claims I made in that chapter.

Are the Findings Obvious?

In referring to that research-based knowledge about teaching effectiveness, it is necessary to deal with the question of obviousness. Are the research findings "for the most part ... generalizations, which are the stock in trade of everyday life for all sane people" (Conant, 1963, p. 133)?

At least three points seem pertinent. First, even if the findings in verbal form are obvious, they are not universally put into practice. Their nonuniversality in practice is evidenced by their correlations with student achievement and attitude. Universality, or anything close to it, would mean zero or very small variability among teachers in the degree to which they use the practice. And such low variability would preclude any correlations with achievement, because a practice cannot correlate, or co-vary, with another variable unless there is some variability among teachers using the practice. So the correlations provide evidence that, however obvious the findings may seem when stated verbally, they are not universally acted upon in actual teaching practice.

The results of experiments on teaching give rise to a second point. In these experiments, some teachers are given training intended to influence them to adopt teaching practices found promising in correlational studies, while others are not given this training. Here are a few of the results of the experiment by Anderson, Evertson, and Brophy (1979):

Classroom Process Variable	*Treatment Group*		*Control Group*		*p*
	M	*SD*	*M*	*SD*	
Percent of all academic interactions that were selected by ordered turns	72	14	24	19	<.01
Percent of incorrect answers that received terminal feedback rather than sustaining feedback (recommended against)	44	8	62	12	<.01
Percent of all academic interactions that included correct answers	73	5	66	8	.04

And here are a few results of the similar experiment by Good and Grouws (1979):

Classroom Process Variable	*Occurrence*		*p*
	Treatment Group	*Control Group*	
Did the teacher conduct review?	91%	62%	.0097
Did the teacher check homework?	79%	20%	.0001
Did the teacher conduct seatwork?	80%	56%	.004
Was the teacher available to provide immediate help to students during seatwork (next five minutes)?	68%	47%	.02

All of these teaching practices are likely to be regarded as obvious. Regardless of their obviousness, these practices were not used by the teachers in the control, or untrained, groups as frequently (when greater frequency had been suggested by earlier correlational studies to be desirable) as they were used by the teachers in the trained groups. After the fact, what research shows seems obvious. But before the research, teachers do these things less often than do teachers who have been given the knowledge of previous research results and their meaning.

The third point on obviousness stems from research on obviousness. Such research has shown that social science findings are rated as obvious, or highly predictable, both when presented as they actually occurred and when presented as the opposite of what actually occurred. For example, Baratz (1983) found that 81% of a group of Stanford undergraduates thought it was fairly obvious that "Those with higher incomes spend a larger portion of their incomes on automobiles and other durable goods than those with lower incomes." And 90% of a group of similar students thought it was fairly obvious that "Those with lower incomes spend a larger proportion of their incomes on automobiles and other durable goods than those with higher incomes." Both of the two opposite statements of a research finding were considered by a strong majority of the students to be fairly obvious. Yet only one of them was true.

It is not unreasonable to expect the same tendency to react with "obvious" to at least some of the findings of research on teaching. Thus I leave it to the reader to say which is more obvious: the finding that "Primary grade students achieve more in classes where they spend most of their time being taught or supervised by teachers rather than working on their own" or "Primary grade students achieve more in classes where they spend most of their time working on their own rather than being taught or supervised by teachers." (The first version is correct; see Brophy and Good, in press, ms. p. 103.)

Evidence on the Improvability of Teaching

It would be gratifying if the evidence on the improvability of teachers in the sense of getting them to use research-based teaching practices could be drawn from work in regular *preservice* teacher education programs. But I, at least, cannot provide that kind of evidence. Instead, I must refer to experiments with *inservice* teacher education programs. These experiments now

number at least nine, and they have dealt with regular teachers at several grade levels. (In describing them, I use "training" as a shorthand term for inservice teacher education.)

One of these experiments was done by Crawford, Gage, Corno, Stayrook, and Mitman (1978) with 33 third-grade teachers who were randomly assigned to a "minimal" training group, a "maximal" training group, or a "control" group. The teachers in the first two groups received packets of reading matter presenting and discussing 32 recommendations based on the findings of four previous correlational studies of teaching effectiveness. The recommendations dealt with classroom management, instructional methods, and questioning and feedback strategies. The minimal training group received only the five packets (one each week) and a self-administered test. The maximal training group received the same materials and also participated in a series of five weekly meetings with a three-person training staff, including an experienced third-grade teacher. What went on in the classes of all 33 teachers was systematically coded by carefully trained observers during about 16 full school days spread over the school year — before, during, and after the training.

Both of the training methods, the minimal and the maximal, had significant effects on the teachers' behavior on dimensions related to the recommendations. Teachers in both trained groups used many of the recommended teaching practices to a greater degree than the control group used them. But significant differences between trained and untrained groups were not found on all of the recommended practices. The more concrete and specific recommendations were more influenced by the training than the more global and complex recommendations were. Also, the minimal training appeared to be just as effective as that which entailed interaction with the training staff.

A second study was done by Anderson, Evertson, and Brophy (1979), who worked with 27 first-grade teachers and focused on management and instruction in small reading groups. The schools of the teachers were divided at random so that no school contained both a trained and a control-group teacher. Ten teachers received both training and observation; the observation occurred once a week from November to May. Seven teachers received only the training but not the observation, and 10 teachers received no training. The training consisted of a manual and two meetings with the project staff for discussion and answering questions on the manual's content. The observation results again showed significant effects of the training when the trained teachers' in-class behavior was compared with that of the

untrained teachers. Again, the more specific recommendations were better acted upon by the trained teachers. For example, methods of choosing students to answer questions were influenced by the training more than methods of using students as models for their peers.

A third study was done by Stallings, Needels, and Stayrook (1978). In a first phase, 25 teachers of junior and senior high school reading received training in the form of six 2.5-hour small-group workshops that provided the teachers with information about their own teaching and with recommendations, encouragement, and continuing feedback. Observations showed that the teachers changed in the recommended ways on most but not all of the practices in teaching reading. A second phase dealt with 32 teachers and yielded similar results. The treatment, or training in this phase, made use of workshops staffed by professionals and was consequently relatively expensive. Again, the recommendations that were less well followed were those that seemed less practical, less congruent with the teachers' values, and less easily adapted to the teachers' own classroom environment.

In a fourth study of this kind, Good and Grouws (1979) worked with 40 fourth-grade teachers of mathematics. Half of the teachers received a 45-page manual dealing with ways of developing content, of conducting practice and review, of pacing and monitoring, and of holding students responsible. The manual was presented to the trained teachers in a 90-minute training session. The teachers also participated in a second 90-minute workshop at the end of the second week of the school term. Observation of all 40 of the teachers during six class sessions showed that the trained teachers used substantially more of the recommended teaching practices than did the untrained teachers.

In a fifth study, the same team of investigators, Good and Grouws (1981), worked with 40 eighth-grade teachers of mathematics. This time the training took two forms. In one kind of training, the teachers worked as *partners* with the investigators in revising the training materials and ideas to suit their own eighth-grade settings. In the other kind of training, the teachers had no such partnership and were merely given the revised program of recommendations in the usual manual-plus-meeting way. Both groups were found through observation to be following the recommendations to a substantially greater degree than an untrained group of teachers. But the two trained groups did not differ, and the partnership feature apparently did not increase the teachers' use of the training.

In a sixth study, conducted in the 1978-79 school year, this string of successes was broken. Theodore Coladarci and I (Coladarci and Gage, 1984) tried to see whether the minimal training that had worked in the original experiment conducted by Crawford and others would work again. Thus the trained teachers were merely given a training manual and some quizzes to take on their own. The whole training program was conducted by mail. For lack of funds, only four observations were made of each of the 40 teachers, both trained and untrained. Two observations were made before the training, and another two were made after the training. The observations showed that the training was ineffective; there were no consistent differences between the trained and untrained groups of teachers in the degree to which they followed the recommendations presented in the training manual, a manual which was the same as that used to good effect with the similar — that is, minimally trained — group in the experiment by Crawford et al. (1978). We can only conjecture as to the reason for the failure of the training to be effective this time. First, "*both* training and classroom observations were held to a minimum" (p. 550). Furthermore, the school district in the failed experiment served poor black children, for the most part, while the two districts involved in the successful training served middle-class white children. Finally, the unsuccessful experiment was conducted right after the passage of California's Proposition 13, which drastically cut school revenues and made teachers anxious. In any case, this sixth experiment showed that minimal "mail-order" training under disadvantageous conditions may not work; there is a point at which training becomes too weak to overcome the obstacles to teacher change.

In a seventh experiment, Emmer, Sanford, Evertson, Clements, and Martin (1981) worked with 41 teachers in grades 1 through 6. One randomly assigned group (N=23) received a manual and participated in two workshops on research-based classroom management principles and strategies for maintaining good learning environments. The other 18 teachers were divided at random into two equal groups, of which one received the training at mid-year; the other received no training. Observations on 12 occasions during the school year showed that the teachers trained at the beginning of the school year used the recommended strategies significantly more than the control group and also had classes that exhibited higher levels of task engagement and appropriate behavior. But the training had only limited effectiveness for the teachers trained at mid-year.

In an eighth experiment, Emmer, Sanford, Clements, and Mar-

tin (1982) worked with 38 junior high school teachers. The experimental group of 18 teachers received a manual and participated in two workshops at the beginning of the school year. All teachers, including the 20 in the control group, were observed, interviewed, and given questionnaires. The data showed that the training had its intended effects on the teachers in most of the nine areas of classroom management, and the classes of the experimental-group teachers showed the desired kinds of task orientations to a greater degree than those of the control-group teachers.

In a ninth experiment, Borg and Ascione (1982) trained an experimental group of 17 teachers in grades 3 through 6 to use a variety of skills for improving pupil behavior and gave a randomly equivalent control group, consisting of another 17 teachers, an irrelevant (self-concept) training program. The nine-week training program provided modules that the teachers studied, applied in simulated situations, practiced in their own classrooms, and checked with partners by means of audiotapes. Observers looked for the 17 specific behaviors in the classrooms of both the experimental-group teachers and the control-group teachers before and after the training. The observations showed that the training resulted in the desired changes in teaching behavior on 14 of the 17 teaching behavior variables; thus the training was, by and large, effective. The pupils in the classes of the experimental-group teachers did better in showing on-task behavior and in avoiding both mildly and seriously deviant behavior.

The main conclusion is that, in eight out of nine attempts, inservice training *was* fairly effective — not with all teachers and not with all teaching practices, but effective enough to change teachers and improve student achievement, attitudes, or behavior. Even in the unsuccessful Coladarci-Gage experiment, it seemed that the recommended teaching practices, derived from previous correlational studies by others, were valid, because the scores of the teachers for conformity to these recommendations correlated, as shown in Figure 1 (p. 21), .40 with mean class achievement.

The Hawthorne Effect?

Before concluding that these experiments demonstrate that *regular* inservice teacher education along the same lines can be effective, some caveats are in order. First, we have no knowledge of how long the effects of these experimental training programs last; no follow-up studies have been made. Second, all of these

nine training efforts were conducted in the context of research enterprises, or experiments. It is not known whether similar training would be effective in the context of regular, routine staff development in the schools. It may well be that a research undertaking leads to a kind of esprit that makes things work better than they usually do in everyday life. This possibility is that of the Hawthorne Effect. As educators know, that effect may produce results in experiments not because of the content and substance of whatever treatment is being manipulated but because of the artificial social forces — that is, the contrived motivational factors — that may operate in experiments.

This possibility should be considered in the light of four facts. First, three major reviews (Cook, 1967; Diamond, 1974; Adair, 1984) have concluded that the Hawthorne Effect is not particularly frequent or widespread in experiments, and only a few examples of it have been found in industry and education.

Second, Good and Grouws (1979) tried to take account of the possibility of a Hawthorne Effect by giving their control group extra motivation to think more about their mathematics teaching. And their control-group classes did show "marked improvement," but not as much improvement as the experimental-group classes.

Table 4

Teaching Practice Variables: Relationship Between Implementation and Correlation with Achievement

		Level of Correlation of Teaching Practice Variable with Achievement	
		Higher	*Lower*
Significance of Difference Between Experimental and Control Groups of Teachers in Implementation of Teaching Practice	Yes	8	0
	No	0	7

Source: Based on data in Good and Grouws (1979).

Third, it is important to realize that these nine experiments had complex, multifaceted experimental treatments that were manipulated in all-or-none fashion. Hence the effects cannot be attributed to any particular component, or teaching practice, in these treatments. But guesses as to the effectiveness of particular components can be based on the correlations between single teaching practices and outcomes. Thus we arrive at a third basis for evaluating the Hawthorne Effect possibility: the results in two of the experiments show that the better a teaching practice was implemented, the more likely it was to correlate with achievement. As is shown in Table 4, of the 15 teaching practices on which teachers were trained in the Good-Grouws experiment, eight were significantly implemented, and all eight were more highly correlated with class mean achievement gain. Of the seven teaching practices that were not significantly implemented, none was more highly correlated with achievement gain. As shown in Table 5, the same kind of relationship was evident, although not as neatly, in the results obtained by Anderson and her associates.

Table 5

Teaching Practice Variables: Relationship Between Implementation and Correlation with Achievement

		Level of Correlation of Teaching Practice Variable with Achievement	
		Higher	*Lower*
Significance of Difference Between Experimental and Control Groups of Teachers in Implementation of Teaching Practice	Yes	12	2
	No	8	9

Source: Based on data in Anderson, Evertson, and Brophy (1979).

Here, of the 14 practices significantly implemented, 12 were more highly correlated with achievement. Of the 17 practices not significantly implemented, nine (only a bare majority) were also not more highly correlated with achievement. These two sets of results suggest that the treatment did not operate in the wholesale, undifferentiated way that the Hawthorne Effect would bring about. The training affected only *some* of the practices at which the training was aimed, and the affected practices tended to correlate more highly with achievement.

Fourth, we shall get rid of our suspicions that it is the Hawthorne Effect and not the content of the training that produces these effects only when the training becomes so routine as to lose its Hawthorne charm. If we still get effects on teaching practices and student achievement or attitude when that day arrives, we shall know it was *not* the Hawthorne Effect that did the trick.

It may turn out that training as brief and relatively inexpensive as that used in these experiments will not work in run-of-the-mill staff development programs. If so, this kind of training will need to be strengthened in various ways. The problem of discovering and developing more effective ways of equipping teachers with pedagogical knowledge and skill is the same, in principle, as that of discovering more effective ways of teaching.

The Observational-Correlational-Experimental Loop and Teacher Education

In the history of science and technology, it is possible to discern a kind of progression that moves from observation and description, through correlational work, to experimentation. This progression can be seen, for example, in the work on cholesterol, whose latest phases were mentioned above. I first heard about cholesterol as a possible factor in heart disease soon after World War II. It was then beginning to be known that cholesterol had been observed, isolated, defined, and suspected of being a factor in heart disease. Then came correlational studies which showed that higher levels of cholesterol were associated with higher rates of heart attack. It was recommended that people watch their intake of cholesterol. But the recommendation was only half-hearted, because there was no solid evidence that the correlation reflected a causal connection, that is, that higher cholesterol level *caused* greater proneness to heart attack. Then came the massive

$150-million 7- to 10-year experiment cited above. Here is how *Science* (Kolata, 1948) discussed that experiment:

> The reason for the trial . . . is that the institute [the National Heart, Lung, and Blood Institute] has never felt comfortable advocating that people reduce their blood cholesterol without good evidence that cholesterol lowering will reduce their heart disease risk. . . . It could be that cholesterol concentrations are correlated with heart disease risk in humans but do not cause the disease . . . (p. 381).

The results shown in Table 3 are those that confirmed the causal character of the relationship. According to the same Kolata article in *Science*, "These findings have already been hailed by the American Heart Association and are expected to affect profoundly the practice of medicine in this country" (p. 381). Thus the descriptive-correlational-experimental loop has apparently been followed in the case of cholesterol.

Another example of the descriptive-correlational-experimental loop occurred in research on the relationships between farming practices and concentrations of fission products in milk (Hansen, Campbell, Fooks, Mitchell, and Eller, 1964). These authors described observations "that milk from cows consuming only stored feed contained no detectable radioiodine while milk from cows grazing on comparable pastures during the same period contained from 50 to 270 pc per liter of iodine-131" (p. 1). Then they conducted a survey to determine which of the many practices involved in dairy farming might be related to levels of iodine-131 and strontium-90 in milk. Finally they carried out an experiment to ascertain whether modifying these farming practices would change the levels of these two elements in milk. Their experiment showed that the levels of the dangerous elements from atmospherically carried nuclear test debris "could be reduced by a factor of 50 percent or more throughout the entire pasturing season by grazing cows on abundant well-fertilized pastures" (p. 30).

So far as I know, this sequence was first advocated for research on teaching by Nuthall and Church (1973), and it was first called the descriptive-correlational-experimental loop by Rosenshine and Furst (1973). This loop is beginning to be used in the case of some teaching practices and student achievement. Teaching practices have been observed, described, and measured. They have then been correlated with achievement. And most recently they have been "manipulated" in the teacher-training experiments and found to cause better student achievement, attitudes, and behavior.

For teacher education and staff development, the same general program seems to make sense. We have now had surveys of the

practices used in teacher education and staff development programs. We have had theoretical analyses by such writers as Fenstermacher and Berliner (1983), who identified four facets of staff development activities that seemed to them to be potentially important. The four facets were:

1. *Source of the initiative for staff development.* Is it initiated internally, by the teachers who will be involved, or externally, by some higher authorities?
2. *Target processes of the staff development.* Are they intended to effect compliance with laws and regulations? Or are they intended to remedy deficiencies of the teachers involved? Or are they intended to enrich teachers' knowledge and skills?
3. *Size of the group involved.* Is it one teacher? Or a few teachers? Or many teachers? Or is it all the teachers of a school, a school district, or a state?
4. *Process of staff involvement.* Was the staff involved by its own free choice, or was it by mandate?

Given these four facets, we can readily conceive of the different staff development programs that would result from different combinations of the various levels of each facet. At one extreme we might have a program initiated by teachers themselves, aimed at enriching teacher knowledge, to be engaged in by a few teachers as a result of their own free choice. At another extreme we might have a program initiated by a state superintendent, designed to effect compliance with a new law, such as that concerning mainstreaming, and intended to affect all teachers in the state, by mandate from the state Capitol.

These two different staff development programs will obviously have different characteristics. Their processes will need to be described along different dimensions by different research techniques. Those processes will need to be correlated with different measures of their effectiveness and success. And eventually, if the descriptive-correlational-experimental loop is followed, experiments will need to be done in which different staff development programs are tried and evaluated.

We are now in the early stages of the task of acquiring this kind of knowledge about how to educate teachers, at either the inservice or the preservice stages of their careers (see, e.g., Sparks, 1983). Whether the work will ever get done will depend on whether policy makers decide that the investment is worthwhile. Some estimates indicate that in recent years staff development, including university courses for inservice teachers, has cost about $2 billion per year in this country. Divided by our approximately two million elementary and secondary school teachers, this

amounts to an investment of about $1,000 per year per teacher in the effort to improve the teachers we already employ. An investment of about 1% of the $2 billion in the effort to understand and improve staff development would provide staff development research with about $20 million a year. As the years went by, that research could make staff development the source of aid and comfort to teachers that it has never been but ought to be.

III
Improving the Gains

To improve the scientific basis of the art of teaching, we should consider some new paradigms for research on teaching — new at least in relation to the ones already used. In doing so, I shall first go briefly into the meaning of *paradigm*. Second, I shall sketch several paradigms that seem now to be the bases for active programs of research on teaching. Third, I shall go into more detail on what seems to be the most active and influential of these new paradigms: the one consisting of the sociolinguistic-ethnographic approach or, more generally, the qualitative approach. I shall draw upon a recent sketch of the goals, methods, and typical findings of the investigators who use this paradigm. Then I shall attempt to relate the paradigm to the process-product paradigm to which the first two chapters were devoted.

What Are Paradigms?

The idea of research paradigm had been around for many decades before the term became widely used. I first heard the term in the 1930s, when B.F. Skinner's operant conditioning was held up as a paradigm different from I.P. Pavlov's classical conditioning. E.G. Boring used the idea of paradigm in the 1950s as a central concept in his discussions of the *Zeitgeist* in the history of psychology. I wrote a chapter entitled "Paradigms for Research on Teaching" (Gage, 1963) without knowing about the work of Kuhn (1962), whose *Structure of Scientific Revolutions* brought wide attention to the concept of paradigm. After Kuhn, almost no discussion of scientific movements proceeded without some reference to paradigms.

But Kuhn used the concept in many different ways, perhaps a dozen or two. Of all those different usages, the one that I have in mind is the one that regards a paradigm as intrinsically an integrated cluster of *substantive* concepts, variables, and problems attacked with corresponding *methodological* approaches and tools. The associated extrinsic manifestations of a paradigm are several. A paradigm gathers unto itself a community of investigators. By sharing information within itself, this community gives itself intellectual and social support. It tends not to communicate with investigators who follow different paradigms. Citation of others' work is frequent within a paradigm but much less frequent, perhaps nonexistent, across paradigms. Hence the followers of a paradigm tend to have their own journals, scientific societies, and meetings, because the paradigm has won their allegiance to an integrated set of concepts, variables, problems, and methods.

In psychology, which is my own field, it is easy to find paradigms. The psychometric paradigm founded by Charles Spearman, L.L. Thurstone, and others lives today in the Psychometric Society and the journal *Psychometrika*. The operant conditioning paradigm founded by B.F. Skinner manifests itself in the Society for the Experimental Analysis of Behavior, the *Journal of the Experimental Analysis of Behavior*, and the *Journal of Applied Behavior Analysis*. The field of research on teaching has also begun to divide into such sharply defined and purified paradigms, but its focus on what goes on in conventional classrooms — on the phenomena, problems, concepts, and variables to be observed in such classrooms — has thus far preserved some unity. It has an informal organization, named the Invisible College, after one of the distinguishing features of paradigms that have been identified by sociologists of knowledge. The Invisible College for Research on Teaching meets annually, but it still lacks the trappings of its own journal and formal organization. The new Division of Teaching and Teacher Education in the American Educational Research Association may foster research on teaching in its several paradigms.

Apart from their other characteristics, paradigms differ in the competencies they require of the investigator. This is an *ad hominem* consideration, but it seems justifiable to confront this aspect of paradigm choice. The sociologist Robert Merton (1969) has written much to show the way in which scientists are "human." Their behavior is influenced not only by rational, altruistic, objective, and impersonal considerations but also by such factors as personal ambition, prestige, and concern with priority in

discovery. Thus it is defensible to discuss the ways in which the personal competencies of research workers influence their choice of paradigm.

The point is simple. What investigators choose as their substantive and methodological paradigms is determined in large part by the knowledge, skills, and intellectual tools they have acquired. If an investigator is equipped with skills in educational and psychological measurement, statistical analysis, and experimental design, he or she is more likely to adopt paradigms requiring these skills. Similarly, the investigator whose background equipped him or her with skills in economic or political analysis, historical research, ethnography, or aesthetic criticism is more likely to make use of research approaches that entail those skills.

What is striking, however, about this tendency to do research in ways that draw on one's personal competencies is the way in which this factor is ignored in controversies about the relative merits of various paradigms. The arguments go on as if no personal factors were involved. The "scientific" investigator decries the lack of objectivity and generalizability of what the ethnographer does. Similarly, the ethnographer derogates the inadequacy, spurious objectivity, and failure to provide a view from the inside, of the user of tests, statistics, and experimental methods. What neither kind of protagonist seems willing to admit is that choices of paradigms are as much determined by personal competencies as by the intrinsic merits of the way researchers have chosen to conduct their careers.

It would be convincing if "scientifically and quantitatively" oriented investigators would find it possible to argue in favor of what the qualitative investigator holds dear. Similarly, it would be altogether salubrious if ethnographic observers would somehow find a place in their scheme of things for the techniques and approaches of the expert in measurement and statistics. In short, the positions of the protagonists on either side of controversies about paradigms would become more convincing if it could be shown that their positions do not correlate substantially with the skills that the protagonists happened to acquire, for reasons usually unrelated to the strengths or weaknesses of their paradigms, while they were graduate students or subsequently. We ought not to confuse the need of investigators to protect their claims to a contribution with the rational merits of their arguments.

Paradigms for Research On Teaching

Are there paradigms within research on teaching? That is, are there strongholds of concepts and methods inside which huddle together mutually supportive groups of investigators who have something less than unbounded admiration for outsiders, the adherents of other paradigms? The answer is yes. Let us turn now to brief descriptions of some of them.

The Process-Product Paradigm

The first two chapters of this volume centered on the process-product paradigm shown in Figure 2, which models the field of research on teaching along lines set forth by Mitzel (1957). The *presage* variables are, in general, teacher characteristics, such as the teacher's age, sex, training, verbal aptitude, years of experience, knowledge of the subject, teacher-education experiences, educational values, and implicit theories of teaching. The *context* variables refer to the setting in which the teaching takes place: the grade level, the subject matter, the students' ability and initial knowledge of the subject, the class size, the socioeconomic status of the pupils in the class, and so on. The *process* variables refer to what goes on in the classroom: teacher behavior, teaching methods, teaching practices and styles, and also student behavior and teacher-student interaction. The *product* variables are those that

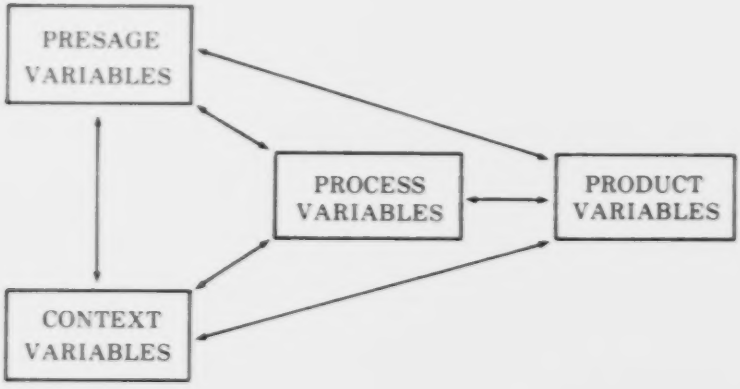

Figure 2. A Model of Research on Teaching (after Mitzel, 1957).

characterize student achievement of cognitive objectives and also student attitudes, feelings, values, and the like. The lines connecting the various pairings of these four kinds of variables have arrows at both ends to suggest that the model implies no commitment to any unidirectionality of influence. For all we know, product variables can affect process variables, or process variables can influence presage or context variables.

The process-product paradigm thus focuses on the relationships between what can be seen and heard in the classroom and what students subsequently exhibit by way of knowledge, understanding, skill, appreciation, and sensibility. Some of those relationships, as I claimed in the first two chapters, are fairly regular, discoverable, and useful in teacher education and the improvement of teaching.

In 1978 Doyle (1978) offered ideas that could be used to extend the process-product paradigm. As is shown in Figure 3, we have

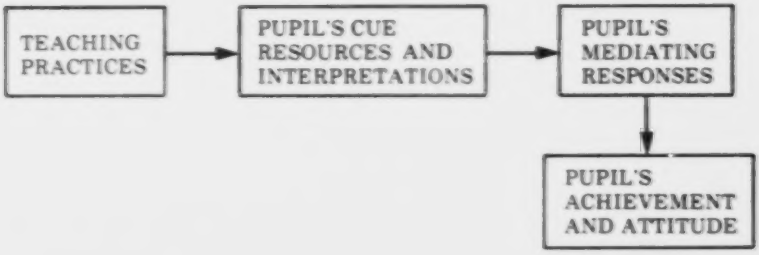

Figure 3. The Extended Process-Product Paradigm

learned that the ecology of the classroom, in the form of cue resources and interpretations, such as tests and praise, guides student behavior. Further, using cues, students make mediating responses, such as attending, translating, segmenting, rehearsing, and otherwise engaging with learning tasks. Thus teacher behavior and teaching practices, as process variables, get filtered through cues and mediating responses so as to influence student achievement and attitude.

The methodology associated with the process-product paradigm comprises typically the structured observation of classroom processes, the achievement test and attitude inventory, the

correlation coefficient in its zero-order and multivariate forms, and the experiment with its random assignment of cases and its manipulation of teaching practices. But the structured observation is not intrinsic to the paradigm, and it is not inconceivable that ethnographic-sociolinguistic observations could be used to advantage in process-product studies. I return to this possibility below.

Other Paradigms

Other paradigms for research on teaching can be identified. Some of these seem to be models for practice rather than paradigms for inquiry. But they qualify as the latter because each of them is associated with a particular set of concepts, variables, and problems around which a research tradition and occasionally some special research methods have developed.

The *behavior modification paradigm* is widely used in the form of token economies for managing the behavior of children and youth, especially those who present problems in the classroom or the special facility. This paradigm focuses on the manipulation of discriminative cues, such as clear sets of rules, and reinforcers, such as credits or tokens that can be traded for privileges of all sorts.

The *interactive educational technology* paradigm uses computers to take the place of the teaching machines that B.F. Skinner devised when he invented programmed instruction. Here again the discriminative cues appear, in the form of sections of prose and questions, to which the student responds. If the response is correct, the student receives a positive reinforcer, in the form of some indication that the student is doing well and may proceed to the next frame. Courses, master's degree programs, textbooks, journals, and societies that have sprung up around this general approach signal its maturing into a full-blown paradigm.

Another group of psychologists is combining behavioral and cognitive approaches into what might be called the *instructional design paradigm* (Reigeluth, 1983). Loosely derived from research and development in military and industrial training, this set of ideas focuses on logically ordered steps and skills considered necessary in designing instructional materials and carrying out instructional programs. The skills take such forms as "analyses of needs, goals, and priorities;" "analyses of course objectives;" "preparation of lesson plans (or modules);" "development or selection of materials and media;" "development of algorithms, heuristics, rule structures, and displays;" "assess-

ment of student performance;" "formative evaluation;" "field testing and revision;" and "summative evaluation." The writings of instructional designers bring to mind not the elementary or secondary school classroom and its teacher but the industrial or business training course for skilled workers in the shop or the office. Their ideas seem to have the connotations of the term "instruction" identified by Eisner (1979), namely, "harsher, less organic, more mechanical than teaching. . . . Instruction is less apt to be associated with what is flexible and emergent . . . and . . . artistic, than is teaching" (pp. 158-159). In any case, instructional design associated with such writers as Robert Gagné, Leslie Briggs, Charles Reigeluth, George Gropper, Lev Landa, David Merrill, and Joseph Scandura does not seem to have been adopted in school teaching or teacher education. Seldom do instructional-design writers give evidence of concern for motivational, emotional, social, disciplinary, or managerial problems of the kind teachers face in elementary and secondary school classrooms. They seem to focus more on cognitive and motor skills and ways of getting trainees to be proficient in them.

Finally, let us return to the *intact teaching styles paradigm* discussed in the first chapter, a paradigm that sets up an integrated set of teaching practices derived from a particular model of learning. One major example of this paradigm is Benjamin Bloom's Mastery Approach. This approach, derived from John Carroll's time-focused model of the factors in school learning, proceeds by analyzing courses or curricula into modules, designing appropriate instructional materials and tests of the mastery of these modules, having alternative instructional materials available to provide corrective feedback for students who do not achieve mastery on the first attempt, and alternative tests to be taken until mastery is achieved. Although it is used in elementary and secondary schools, Bloom's Mastery Approach is hard to distinguish from Fred Keller's Personalized System of Instruction, used at the college level. The latter is another intact teaching style, one derived, according to Keller, from Skinnerian principles. Keller's paradigm has a journal, a society, annual meetings, and a substantial, mutually supportive corps of adherents.

This multiplicity of paradigms in research on teaching may strike some observers as symptomatic of the immaturity of the enterprise. But such multiplicity is found in all scientific fields. As Dewey (1928) noted:

> [T]here is nothing false or extravagant in declaring that at the present time different sciences of education are not only possible but also much needed. Of course such a statement goes contrary to

> the idea that science by its very nature is a single and universal system of truths. But this idea need not frighten us. Even in the advanced sciences, like those of mathematics and physics, advance is made by entertaining different points of view and hypotheses, and working upon different theories. The sciences present no fixed and closed orthodoxy (p. 116).

In any case, there are several active paradigms for research and development in teaching other than the process-product paradigm. But none of the other paradigms seems to be primarily concerned with what is going on in the vast majority of the classrooms in American elementary and secondary schools. Whatever their strengths and victories, they represent radical departures from what we might call traditional-conventional classroom teaching.

The Persistence of Classroom Teaching

Many movers and shakers have tried to replace classroom teaching. But it persists nonetheless, after at least a century of dominance, in the schools not only of America but of most other countries on which we have data.

The persistence of teacher-centered classroom teaching has been studied as a historical phenomenon by Cuban (1984). He documented what he called "a seemingly stubborn continuity in the character of instruction" (p. 21), one that has persisted despite attempts in this century to introduce the Dalton Plan, the project method, team teaching, the open classroom, and so on.

By "teacher-centered instruction," Cuban meant what I have called classroom teaching, namely, teaching where

> —Teacher talk exceeds student talk during instruction.
> —Instruction occurs frequently with the whole class; small-group or individual instruction occurs less frequently.
> —Use of class time is determined by the teacher.
> —The classroom is usually arranged into rows of desks or chairs facing a blackboard with a teacher's desk nearby (p. 3).

Cuban looked at the evidence about classrooms available around the turn of the century, during the 1920s and 1930s, and the years since the mid-1960s, in several large cities and in rural areas, in many of which there had been strong reform movements. He looked at teacher-written accounts, newspaper articles, principals' reports, student recollections, photographs, student publications, formal studies and surveys of school systems, and the responses of teachers and students to questionnaires on

classroom activities. What Cuban found was that classroom practices tended to be stable; they resembled the approaches used by former teachers. About 90% of high school teachers continued to use traditional teacher-centered methods, and perhaps 75% of the teachers in the elementary schools did the same. It seems, therefore, that only a small minority of teachers adopted any of the practices associated with progressive or open education.

Cuban and others (e.g., Westbury, 1973) have tried to account for the impregnability of traditional classroom teaching. For our present purposes, however, this persistence suggests that the new paradigms of teaching — and of research on teaching — mentioned above are unlikely to take hold. They amount to trying to change the basic art form that constitutes classroom teaching. They are tantamount to trying to persuade poets to stop writing sonnets and start writing sonatas. Cuban's analysis of the causes of this stability of the art form in teaching suggests that not only open education but other radical departures from traditional classroom teaching are likely to be ephemeral.

My analogy with art forms seems to apply to the approach to research and development in teaching and teacher education that I see as most promising. If we cannot replace the art form, we can perhaps change the quality of what is done in that form. Within the sonnet form there are poor sonnets and good sonnets. The same is true of concertos, watercolors, and mobiles. Within the art form called classroom teaching there is also great variation in quality. What I see as promising is research that accepts the basic parameters of classroom teaching: teacher-centeredness, whole-class organization, subject-matter orientation, and much recitation, interspersed with short lectures, discussions, tutoring, and seatwork. Such research will study the infinite variations possible within this stable art form, the variations that make the difference between superb, average, and atrocious classroom teaching.

All this brings me back to the process-product paradigm for research on classroom teaching. As the reader will recall, I have claimed (and referred to documentation to support my claim) that such research has already yielded successes in changing teaching practices, *within the classroom teaching mode*, in ways that *cause* substantially improved student achievement, attitude, and behavior. The successes thus far have been attained primarily in the teaching of reading and mathematics in the early grades, but not only in those areas. And the descriptive-correlational-experimental research sequence that has yielded these successes can be used to obtain similar improvements at other grade levels,

in other subject-matters, in other systems of educational values, and in inservice and preservice teacher education.

The Technological-Deliberative Distinction

The implications of what I have been saying have raised questions. Zumwalt (1982) held that process-product researchers take a "technological" orientation toward teaching and teacher education as opposed to a "deliberative" orientation. A "deliberative orientation" is one in which "concern is not that teachers follow a set of rules, which could never account for all circumstances anyway, but rather that teachers view teaching as a process of constantly making choices about means and ends — choices that can be informed by process-product research, descriptive research, experience, intuition, and one's own values" (p. 226).

What I (1978) have emphasized in speaking of an art of teaching shows that process-product research is not tied to a technological approach. There seems to me to be ample room for departures from process-product findings in my characterization of teaching as "a process that calls for intuition, creativity, improvisation, and expressiveness — a process that leaves room for departures from what is implied by rules, formulas, and algorithms." I refer also to my statements that in teaching "there is a need for artistry: in the choice and use of motivational devices, clarifying definitions and examples, pace, redundancy, and the like," and that "no one can ever prescribe successfully all the twists and turns to be taken as the lecturer, the discussion leader, or the classroom teacher uses judgment, sudden insight, sensitivity, and agility to promote learning" (Gage, 1978, p. 15).

Similarly, the equating of process-product research with a technological, or nondeliberative, approach to teacher education is belied by Smith (1980), whom Zumwalt also classified as a member of the technological-orientation camp. Thus Smith wrote that clinical principles of teaching can be learned verbally or as ways of behaving. He said:

> Both modes of learning are necessary in a profession. Just to learn them verbally is to render the principles useless in practice. To follow them behaviorally *without understanding* is to use them mindlessly. . . . [T]he disadvantage of performance without conceptual understanding [is that it puts teachers] at the mercy of their own propensities. . . . Reasoning from case to case without mediating principles is groundless and points up both the necessity for teaching principles per se and for using them to mediate skillful performance in professional training (pp. 65-66; italics added).

Note Smith's emphasis on the teacher's *understanding* of principles as a basis for using them to *mediate* performance.

In my view Zumwalt's conception of a technological orientation — holding that the implications of process-product research should be followed exclusively or nondeliberatively — fits few if any process-product research workers. It is true that process-product researchers have given teachers recommendations for teaching practices derived from process-product correlations. But that has been done solely within an experimental-research context. These researchers have never expected teachers to adopt nondeliberatively all the recommended practices under all circumstances. Indeed, the results of these experiments, mentioned above, have shown that not all the teachers did adopt all the practices. The researchers have not regarded decisions on the part of some teachers not to use some of the recommended practices as irrational, erroneous, or unjustified. Process-product researchers also insist "that research data can be *one*, but not the only, input in the deliberative process" (Zumwalt, 1982, p. 227). Zumwalt advocated "an emphasis on developing the reasoning powers of teachers *rather than* providing them with substantive findings to implement" (p. 239, italics added), apparently without realizing that the substantive findings can give teachers something to reason with and about.

The Complementary Uses of Central Tendency and Variability

Many controversies similar to this one — about the nature of teaching and the value of scientific research — relate to the universal distinction between the central tendency and the variation around that central tendency. That distinction shows up in many contexts: Social roles represent central tendencies, and unique personalities represent variation around roles. Classrooms possess many uniform characteristics, yet each classroom is in many ways unique. Any individual teacher's use of a given teaching method, such as classroom teaching, hovers around a central tendency and also varies away from that central tendency.

For this reason, when research reveals correlations between variables, we can value the correlations as revealing central tendencies in the relationship between the variables. Or we can disparage the correlations because they neglect to tell us about the departures from the regression line. This conflict between emphasis on the regular (the central tendency) and emphasis on the

particularistic (the variation) can be found within the same writer — even within the same article. Thus one writer said early in his paper that the teacher's pedagogical "knowledge derives not so much from a systematic comparison of a number of similar situations as it does from an intuitive analysis for a specific context in which *many important qualities are assumed to be unique*" (Bolster, 1983, p. 298, italics added). Later, this writer said that "teachers' expectations and techniques, to be credible, must be acceptable variants of *a general pattern of professional behavior* called 'teaching' "(p. 298, italics added) and that "as in all professional behavior, teaching tends to become institutionalized . . . and it engenders a set of widely held expectations that broadly define appropriate teacher behavior" (p. 300). Note that this author first emphasized the *specificity and uniqueness* of each classroom and later emphasized the *general pattern* and *widely held expectations*. He was right in both instances. But he can be interpreted as disagreeing with himself.

When different writers adopt one of these different emphases, they easily foment controversy. But the controversy is as spurious as that concerning whether a glass is half full or half empty. The supposed disjunction between the so-called technological and deliberative orientations has this character. Of course, it is valuable to have central tendencies, or generalizations — even the weak ones that empirical research sometimes yields in education and medicine. Such generalizations, like all central tendencies, give us a basis for thinking about broad trends and likelihoods. But of course it is also desirable to remain eternally aware that cases vary around averages. Thus "personality" makes people in the same "role" behave differently to some extent. Thus no specific third-grade teaching of reading need yield exactly the product that the correlations lead us to expect from a whole sample of instances of third-grade teaching of reading.

The wise researcher, teacher educator, or teacher will use the central tendencies for all they are worth. And some use of trends is, as Bolster (1983) pointed out, well-nigh inevitable if a teacher is to make sense out of the bewildering infinitude of events in the classroom. But the same wise person will also be sensitive to the uniqueness of every classroom and student and alert to the likelihood of departures from the central tendencies that past experience or research may have revealed.

The Ethnographic-Sociolinguistic Paradigm

Arguing for the process-product paradigm does not diminish the value of other kinds of research that can enlighten teachers' deliberations. Thus we should also consider the value of ethnographic-sociolinguistic research for the enterprise of understanding and improving teaching. Let us turn now to that paradigm. In doing so, I shall draw on a succinct and thoughtful characterization of such research by Bolster (1983). Ethnographic-sociolinguistic research views teaching as an instance of symbolic interaction, the process whereby "people act on the basis of the meanings events have for them" (p. 303). The meanings arise out of social interaction. Persons develop and change these meanings in terms of the situation and their own actions. Significant knowledge about any situation, including the classroom, consists of awareness of the meanings that participants develop about that situation. The purpose of the corresponding research is to get "systematic and reliable information which teachers can use to shed light on their own pedagogical situations" (p. 304).

Such research seeks "thick, critical descriptions" of what is going on in classrooms. The descriptions are full of concrete detail to which interpretations (pointing to the significant patterns of phenomena) and explanatory propositions are attached. The ethnographic methodology is qualitative and systematic. The explanatory propositions are developed over time as the researchers constantly juxtapose salient patterns of events and the meanings that emerge. The interpretations are validated against their consistency with recurrences of the events and also against the meanings that teachers and students attach to the classroom situation.

What kinds of phenomena does such work reveal? Recent reviews (Cazden, in press; Green and Smith, 1983) show that ethnographic-sociolinguistic research has dealt with such aspects of classroom discourse as participation, lesson components, tentativeness indicators, peer talk around academic tasks, mismatches of form and function in speech, and the talk used by teachers to control classroom events. The reports describe phenomena that are sometimes already known and sometimes not already known (such as a surprising rule for showing respect) and also sometimes welcome and sometimes unwelcome (as when teachers are shown to lack influence on something educationally important) (Hymes, 1982, pp. 1-5).

So far, so good. It is hard to imagine that anyone could fail to value such work. Although I have never done such observation, I applaud those who do it.

But then Bolster stated that the "basic stance" of such research is "idiographic"; that is, it aims at understanding "a particular classroom rather than the experimental derivation of a selected number of elements whose relationship can be replicated elsewhere" (p. 305). That proviso gives me pause. It implies that what is learned ethnographically (hence laboriously and expensively) about one classroom will have no value in contributing toward a central tendency for use in understanding other classrooms. With two million teachers, we shall never understand our classrooms, even this year, to say nothing of next year, when two million new classrooms, each presumably unique, will come into being. Does every classroom therefore need its own ethnography if the benefits of such research are to be offered to every teacher? Can teachers ever adequately realize the benefits that Bolster has discerned, since he says that teachers find the experience of discussing the ethnography of their own class "both professionally productive and intellectually stimulating" and also that teachers "develop more differentiated and reflective views of their teaching" (p. 306)?

No. Bolster eventually must face the limitations imposed by his idiographic position. One limitation is that the findings are therefore "context specific" (p. 307), and "inquiries of this type have no demonstrable generalizability" (p. 307). This tremendous limitation seems to me to be potentially fatal to the whole idea of ethnographic research. Bolster seemed to apprehend the same dire implication, because he was forced to retreat from his insistence on total context specificity. Thus he said:

> Granting the importance of contextual variables to specific outcomes, it seems reasonable to assume that *there are broad areas of similarity among classroom environments*, especially at the same grade levels. Consequently, when a sizeable body of ethnomethodological descriptions accumulate, we will have strong indications of how teaching effects occur in various types of setting. It may appear paradoxical to claim that what ethnographers perceive as significant are the particularistic qualities of the classroom but also that such research may spawn universal propositions about teaching. On the other hand, it would be simplistic to argue [as I think Bolster did argue earlier in his paper] that all classroom variables are inherently unique to their particular setting or have no significance except in that specific configuration. Thus qualitative methodologies should be especially useful in research on teaching, since they allow for specifying and elaborating the nature of relationships that hold across settings (p. 307; italics added).

In short, ethnographic-sociolinguistic work can eventually be used for the same nomothetic purposes — abstraction-describing, generalization-seeking, central-tendency finding — that other forms of observation and description are used for.

Thus far, ethnographic work has been primarily descriptive, contributing to our understanding of the way teaching is. Can such research go beyond providing the "anthropological explanations of school failure" that Cazden (1983) saw as a plausible (even if regrettable) characterization of the yield of ethnographic research? In short, Cazden asked, "Can ethnographic research go beyond the status quo?" In response to her own question, she cited the work of an ethnographer (Shirley Brice Heath) who worked with teachers to design and try out "new curriculum and new patterns of classroom discourse" (p. 35). She also cited a team project — the Kamehameha Early Education Project — in which ethnographers, psychologists, and educators developed over a 10-year period "a program that has had dramatic benefit for Polynesian children's achievements in reading" (p.35).

If ethnographers are motivated to seek generalizations and also to foster the improvement and not merely the understanding of teaching, it becomes conceivable that they will collaborate with the investigators who have been pursuing process-product research of the kind that yields implications for desirable changes in classroom teaching practice. The ethnographer's role in such collaboration will be, at the least, to contribute more thorough and subtle descriptions of classroom process. The process-product researcher's role will be to analyze the relationships between such ethnographically developed aspects of process and the best possible measures of classroom "product" in the form of student achievement, attitude, and behavior.

Ethnographers have unquestionably gone deeper into classroom processes, in more detail, with greater sophistication, than any of their predecessors in classroom observation. One need only compare an ethnographic report with what one gets from observations with the categories developed by Ned Flanders to appreciate how far classroom observation has come during the last quarter century. But the greater depth of ethnography and sociolinguistic research does not mean that the divorce of the ethnographic paradigm from the extended process-product paradigm must be continued.

As ethnography yields new phenomena and as replications determine the generalizability of those phenomena, in ways that Bolster (1983) hoped for, there is no requirement that we continue to forgo what statistical analyses and quantitative methods have

to offer. We can exploit the power of both "numbers and narrative," which Light and Pillemer (1984) saw as useful for the problem of research synthesis. A single ethnography can prove that a certain phenomenon is *possible*. Only a number of ethnographies can tell us how frequent and *probable* that phenomenon seems to be. Given a number of ethnographies along with data collected in other ways — as with questionnaires or even, although the thought is unpopular in some places these days, with tests — we can begin to learn about the relationships between ethnographically revealed phenomena and other kinds of phenomena. Those other kinds of phenomena can include measures of student achievement and attitude. Perhaps we can also look at process-product relationships more analytically, relating classroom process to the achievement of a single student on a single test question instead of the achievement of a whole class on a whole test.

In short, we need to see some connection between the sociolinguistic paradigm and the process-product paradigm. I hope the connection leads to a betrothal. Indeed, let us not to the marriage of true paradigms admit impediments.

A Time Perspective

How long will it take for research on teaching to provide a substantial scientific basis for the art of teaching and for the education of teachers in that art? Some writers are ready to give up the effort. They begin the history of the effort with Joseph Mayer Rice in the 1890s and trace it through E. L. Thorndike in the early 1900s, Arvil S. Barr in the 1930s and '40s, and the dozens of workers who undertook research on teaching in the 1960s and '70s. They conclude that, despite much effort, the research basis for teacher education remains weak and is a futile enterprise.

What such writers disregard is the difference in adequacy between early research on teaching effectiveness and research done since the mid-1960s. Before then, relatively few major research programs were undertaken. Most research until the mid-1960s consisted of master's and doctor's dissertations. Only rarely was the research done by persons devoting a substantial part of their careers to such work. Only rarely was the research based on observations of classroom processes, because such observations cost too much for the typical dissertation. Personality tests and rating scales predominated as measures of teachers and teaching.

Ratings by principals and student teaching supervisors predominated as measures of teaching effectiveness.

With the advent of federal support for educational research, the picture began to change in the middle 1960s. The number of long-term workers engaged in cumulative programs of research on teaching increased somewhat, although not nearly enough. The sophistication of research methodology in observation, measurement, statistics, and theory improved. A considerable body of research findings accumulated. Methods of integrating those findings also improved substantially.

Whether the present research basis is weak has been considered above. Here we note only that the effort so far has not been great in proportion to the importance and complexity of the problem. Other research problems, such as the cause and cure of schizophrenia or melanoma, have been pursued over the centuries without success. Yet the lack of success has not caused cessation of the effort to conquer those diseases. Still other diseases, after similar centuries of failure, have finally yielded to persistent research effort: Hodgkin's disease, detached retinas, and the host of bacterial diseases conquered by penicillin, for example.

The argument of the foregoing pages has been that the research effort aimed at a scientific basis for the art of teaching is worth continuing. But how long will it take to build an adequate scientific base? Premature exploitation of inadequate bases can lead to hostility and despair, as can be seen in the efforts to institutionalize performance-based teacher education in the late 1960s and early 1970s (Gage and Winne, 1975). We should avoid such grand but premature applications, just as in aviation we avoided development of national airlines during the years between the Wright brothers' first flight in 1903 and the early 1930s, when aircraft became dependable. We are now, perhaps, at a stage comparable to that of 1913 in the history of aviation.

The question of how long it will take to get a major payoff from research on teaching cannot be answered at present with John F. Kennedy's precision in foretelling in 1961 that an American would walk on the moon by 1969. Nonetheless, guesses can be offered. For those who accept the values embodied in present-day tests in elementary school reading, mathematics, social studies, and science, it is already possible to point to classroom management practices that will improve student achievement, attitude, and conduct. Careful use of the scientific basis already established for the artistic practice of classroom management in elementary school grade levels is highly likely even now to result in an improvement of teacher education, both preservice and inservice. It is also highly likely to result in appreciable improve-

ment in what students learn in these school subjects, how they feel about those subjects and themselves, and how much on-task behavior they show.

The risks in applying the currently available scientific basis seem relatively low, while the probable gains seem fairly substantial. The cost in time, effort, and money also seems low. The psychological costs in disruption of established patterns of school organization, curriculum development, and teaching practice seem low, because the scientific basis arises from observation of what some present-day teachers — the ones who are more effective in improving their students' achievement and attitudes — have already been doing. Influencing teachers to adopt these more effective practices has required no major investments or disruptions. At least in research contexts, the influence has been effective even though teachers spent only a few hours a month in reading and attending meetings.

What remains to be done in the classroom management domain is to try the research-based teaching practices and teacher education programs outside of research contexts — in regular preservice and inservice teacher education programs. The preservice effort will call for careful examination of the experiments cited above to work out ways in which the content and methods of those experiments can be adapted to the formats that work in teacher education institutions. The inservice effort will call for programs conducted by teachers' organizations and school districts. In all these efforts, both preservice and inservice, teachers should be given the full story: how the research-based practices were identified, why they seem reasonable, how they work, and what questions might be raised about their scientific and moral bases. Each teacher should be asked to confront the moral issue of whether these practices can justifiably be rejected on the basis of the teacher's own experience, intuitions, hunches, or predilections. If teachers can justify a rejection of these findings on good intellectual and moral grounds, they should do so. But if they cannot, then they should face the kinds of moral forces that led them to become teachers in the first place — forces that led them to seek to help others achieve personal goals and social values.

In my view, continued effort should be based on a marriage of ethnographic-sociolinguistic observational approaches and the process-product paradigm's observational-correlational-experimental loop. Such effort could lead in a decade or two to a basic strengthening of what goes on in teacher education and in classrooms. This prospect assumes no magical breakthroughs. Rather, it calls for the same kind of steady effort that has im-

proved cures for Hodgkin's disease, surgery for detached retinas, social roles for women, freedom from racial discrimination in employment, equality of educational opportunity, and human longevity. The natural and social sciences have contributed magnificently to these advancements of the human condition. Over the decades, scientific method can do the same for teaching and teacher education.

References

Note: Numbers in brackets at the end of a reference indicate the pages in the text on which the reference is cited.

Adair, J. G. "The Hawthorne Effect: A Reconsideration of the Methodological Artifact," *Journal of Applied Psychology*, March 1984, pp. 334-45. [35]

Anderson, L. M.; C. M. Evertson; and J. E. Brophy. "An Experimental Study of Effective Teaching in First-Grade Reading Groups," *Elementary School Journal*, March 1979, pp. 193-223. [23, 28-29, 31-32, 36]

Axelrod, J. *The University Teacher as Artist*. San Francisco: Jossey-Bass, 1973. [5]

Baratz, D. "How Justified Is the 'Obvious' Reaction?" Doctoral dissertation, Stanford University, 1983. [30]

Billington, D. P. *The Tower and the Bridge: The New Art of Structural Engineering*. New York: Basic Books, 1984. [8]

Block, J. H. and R. B. Burns. "Mastery Learning." In L. S. Shulman, ed., *Review of Research in Education*. Itasca, Ill.: F. E. Peacock, 1976. [18]

Bolster, A. S. "Toward a More Effective Model of Research on Teaching," *Harvard Educational Review*, August 1983, pp.294-308. [52, 53, 54]

Borg, W. R. and F. R. Ascione. "Classroom Management in Elementary Mainstreaming Classrooms," *Journal of Educational Psychology*, February 1982, pp. 85-95. [23, 34]

Brophy, J. E. and T. L. Good. "Teacher Behavior and Student Achievement." In M. E. Wittrock, ed., *Handbook of Research on Teaching*, 3rd ed. New York: Macmillan, forthcoming. [20, 30]

Cazden, C. "Can Ethnographic Research Go Beyond the Status Quo?" *Anthropology and Education Quarterly*, Spring 1983, pp. 33-41. [55]

_____. "Classroom Discourse." In M. E. Wittrock, ed., *Handbook of Research on Teaching*, 3rd ed. New York: Macmillan, forthcoming. [53]

Coladarci, T. and N. L. Gage. "Effects of a Minimal Intervention on Teacher Behavior and Student Achievement," *American Educational Research Journal*, Fall 1984, pp. 539-55. [23, 33, 34]

Conant, J. B. *The Education of American Teachers*. New York: McGraw-Hill, 1963. [27, 28]

Cook, D. L. *The Impact of the Hawthorne Effect in Experimental Designs in Educational Research: Final Report*. Columbus: Ohio State University, 1967. [35]

Crawford, J. and N. L. Gage. "Developing a Research-based Teacher Education Program," *California Journal of Teacher Education*, vol. 4, 1977, pp. 105-123. [20]

Crawford, J.; N. L. Gage; L. Corno; N. Stayrook; and A. Mitman. *An Experiment on Teacher Effectiveness and Parent-assisted Instruction in the Third Grade*, vols. 1-3. Stanford, Calif.: Program on Teaching Effectiveness, Center for Educational Research at Stanford, 1978. (ERIC Document Reproduction Service No. ED 160 648) [23, 31, 33]

Cronbach, L. J. "Beyond the Two Disciplines of Scientific Psychology," *American Psychologist*, February 1975, pp. 116-27. [9]

Cuban, L. *How Teachers Taught: Constancy and Change in American Classrooms* 1890-1980. New York: Longman, 1984. [3, 11, 48-49]

Dewey, J. "Progressive Education and the Science of Education," *Progressive Education*, July-September 1928, pp. 197-204. Reprinted in M. S. Dworkin, ed., *Dewey on Education: Selections with an Introduction and Notes*. New York: Bureau of Publications, Teachers College, Columbia University, 1959. [47-48]

Diamond, S. S. *Hawthorne Effects: Another Look*. Chicago: The Author, University of Illinois, 1974. [35]

Doyle, W. "Paradigms for Research on Teacher Effectiveness." In L. S. Shulman, ed., *Review of Research in Education* (pp. 163-198). Itasca, Il.: F. E. Peacock, 1978. [45]

Eisner, E. W. *The Educational Imagination: On the Design and Evaluation of School Programs*. New York: Macmillan, 1979. [47]

Emmer, E. T.; J. P. Sanford; B. S. Clements; and J. Martin. *Improving Classroom Management and Organization in Junior High Schools: An Experimental Investigation*. Austin: Research and Development Center for Teacher Education, University of Texas, 1982. [23, 33]

Emmer, E. T.; J. P. Sanford; C. M. Evertson; B. S. Clements; and

J. Martin. *The Classroom Management Improvement Study: An Experiment in Elementary School Classrooms*. Austin: Research and Development Center for Teacher Education, University of Texas, 1981. (ERIC Document Reproduction Service No. ED 178 460) [23, 33]

Eysenck, H. J. "Meta-Analysis: An Abuse of Research Integration." In J. A. Wakefield, ed., *Application of Personality Theory to Education*, forthcoming. [15]

Fenstermacher, G. and D. C. Berliner. *A Conceptual Framework for the Analysis of Staff Development*. Santa Monica, Calif.: Rand Corporation, 1983. [39]

Gage, N. L. "Paradigms for Research on Teaching." In N. L. Gage, ed., *Handbook of Research on Teaching*. Chicago: Rand McNally, 1963. [41]

_____. "Four Cheers for Research on Teaching." Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, 1976. [15]

_____. *The Scientific Basis of the Art of Teaching*. New York: Teachers College Press, 1978. [15, 16, 50]

_____. "When Does Research on Teaching Yield Implications for Practice?" *Elementary School Journal*, March 1983, pp. 492-96. [14]

Gage, N. L. and T. Coladarci. *Replication of an Experiment with a Field-based In-Service Teacher Education Program: Final Report*. Stanford, Calif.: Center for Educational Research at Stanford, 1980. (ERIC Document Reproduction Service No. ED 181 023) [20, 21]

Gage, N. L. and R. Giaconia. "Teaching Practices and Student Achievement: Causal Connections," *New York University Education Quarterly*, Spring 1981, pp. 2-9. [23]

Gage, N. L. and P. H. Winne. "Performance-based Teacher Education." In K. Ryan, ed., *Teacher Education: 74th Yearbook of the National Society for the Study of Education, Part 2* (pp. 146-72). Chicago: University of Chicago Press, 1975. [57]

Gewirth, A. "Can Man Change Laws of Social Science?" *Philosophy of Science*, July 1954, pp. 229-41. [10]

Giaconia, R. and L. V. Hedges. "Identifying Features of Effective Open Education," *Review of Educational Research*, Winter 1982, pp. 579-602. [17-18]

Glass, G. V. "Primary, Secondary and Meta-Analysis of Research," *Educational Researcher*, November 1976, pp. 3-8. [15]

Good, T. L. and D. A. Grouws. "The Missouri Mathematics Effectiveness Project: An Experimental Study in Fourth-Grade Classrooms," *Journal of Educational Psychology*, June 1979, pp. 355-62. [23, 29, 32, 35, 36]

Good, T. L. and D. A. Grouws. *Experimental Research in Secondary Mathematics Classrooms: Working with Teachers.* (NIE-G-79-0103 Final Report) Columbia, Mo.: The Authors, University of Missouri, 1981. (ERIC Document Reproduction Service No. ED 219 261) [23, 32]

Green, J. L. and D. Smith. "Teaching and Learning: A Linguistic Perspective," *Elementary School Journal*, March 1983, pp. 353-91. [53]

Hagberg, H. M. "Don't be Teachin' Jovita Wrong: An Exploration of Reading Instruction in Four Urban Third-Grade Classrooms." Doctoral dissertation, Stanford University, 1983. [27]

Hansen, W. G.; J. E. Campbell; J. H. Fooks; H. C. Mitchell; and C. H. Eller. *Farming Practices and Concentrations of Fission Products in Milk.* Washington, D. C.: Division of Radiological Health, Public Health Service, U. S. Department of Health, Education, and Welfare, 1964. [38]

Hativa, N. "Factors of Organization and Clarity in Mathematics Lessons." Doctoral dissertation, Stanford University, 1981. [27]

Hedges, L. V. and I. Olkin. "Vote-Counting Methods in Research Synthesis," *Psychological Bulletin*, September 1980, pp. 359-69. [15]

Highet, G. *The Art of Teaching.* New York: Vintage Books, 1955. [4]

Horwitz, R. "Psychological Effects of the Open Classroom," *Review of Educational Research*, Winter 1979, pp. 71-86. [16]

Howe, H. II. "Improving Teaching: Solutions Close to the Classroom," *College Board Review*, Fall 1982, pp. 12-15, 28-29. [26]

Hymes, D. *Ethnolinguistic Study of Classroom Discourse.* Final report to the National Institute of Education, 1982. (ERIC Document Reproduction Service No. ED 217 710). [53]

Jaynes, J. "Hello, Teacher: A Review of F. S. Keller, *The History of Psychology: A Personalized System of Instruction*, and F. S. Keller, *Selected Readings in the History of Psychology: A PSI Companion*; *Contemporary Psychology*, August 1975, pp. 629-31. [19]

Kennedy, M. M. "Findings from the Follow Through Planned Variation Study," *Educational Researcher*, June 1978, pp. 3-11. [18]

Kolata, G. "Lowered Cholesterol Decreases Heart Disease," *Science*, January 27, 1984, pp. 381, 382. [13, 38-39]

Kuhn, T. S. *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press, 1962. [41]

Kulik, J. A.; C. C. Kulik; and P. A. Cohen. "A Meta-Analysis of

Outcome Studies of Keller's Personalized System of Instruction," *American Psychologist*, April 1979, pp. 307-18. [18]

Levin, H. M. *Cost-Effectiveness: A Primer*. Beverly Hills, Calif.: Sage Publications, 1983. [14]

Light, R. J. and D. B. Pillemer. *Summing Up: The Science of Reviewing Research*. Cambridge, Mass.: Harvard University Press, 1984. [56]

Merton, R. K. "Behavior Patterns of Scientists," *American Scientist*, Spring 1969, pp. 1-23. [42]

Mitzel, H. E. "A Behavioral Approach to the Assessment of Teacher Effectiveness." Unpublished paper, Division of Teacher Education, College of the City of New York, 1957. [44]

Nuthall, G. and J. Church. "Experimental Studies of Teaching Behaviour." In G. Chanan, ed., *Towards a Science of Teaching* (pp. 9-25). Windsor, UK: NFER Publishing Co., 1973. [38]

Peterson, P. L. "Direct Instruction Reconsidered." In P. L. Peterson and H. J. Walberg, eds., *Research on Teaching*. Berkeley: McCutchan, 1979. [16]

Reigeluth, C., ed. *Instructional Design Theories and Models: An Overview of Their Current Status*. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1983. [46]

Rosenshine, B. and N. Furst. "The Use of Direct Observation to Study Teaching." In R. M. W. Travers, ed., *Second Handbook of Research on Teaching* (pp. 122-83). Chicago: Rand McNally, 1973. [38]

Rosenthal, R. and R. L. Rosnow. *The Volunteer Subject*. New York: Wiley, 1975. [22]

Rosenthal, R. and D. B. Rubin. "A Note on Percent Variance Explained as a Measure of the Importance of Effects," *Journal of Applied Social Psychology*, September-October 1979, pp. 395-396. [11]

Schön, D. A. *The Reflective Practitioner: How Professionals Think in Practice*. New York: Basic Books, 1983. [7]

Shavelson, R. J. and P. Stern. "Research on Teacher's Pedagogical Thoughts, Judgments, Decisions, and Behavior," *Review of Educational Research*, Winter 1981, pp. 455-95. [6]

Smith, B. O., in collaboration with S. H. Silverman, J. M. Borg, and B. V. Fry. *A Design for a School of Pedagogy*. Washington, D. C.: U. S. Government Printing Office, 1980. [50-51]

Smith, M. L. and Gene V Glass. "Meta-Analysis of Psychotherapy Outcome Studies," *American Psychologist*, September 1977, pp. 752-60. [9]

Sparks, G. M. "Inservice Education: Training Activities, Teacher Attitude, and Behavior Change." Doctoral dissertation, Stanford University, 1983. [39]

Stallings, J.; M. Needels; and N. Stayrook. *How to Change the Process of Teaching Basic Reading Skills in Secondary Schools: Phase II and Phase III*. Menlo Park, Calif.: SRI International, 1978. [23, 32]

Stedman. L. C. and M. S. Smith. "Recent Reform Proposals for American Education," *Contemporary Education Review*, Fall 1983, pp. 85-104. [26]

Tom, A. *Teaching as a Moral Craft*. New York: Longman, 1984. [5]

Westbury, I. "Conventional Classrooms, 'Open' Classrooms, and the Technology of Teaching," *Journal of Curriculum Studies*, May 1973, pp. 99-121. [49]

White, K. R. "The Relation Between Socioeconomic Status and Academic Achievement," *Psychological Bulletin*, May 1982, pp. 461-81. [9]

Wissot, J. "Teaching as an Art Form," *Educational Forum*, March 1979, pp. 265-78. [5]

Woolfolk, A. E. and D. M. Brooks. "Nonverbal Communication in Teaching." In E. W. Gordon, ed., *Review of Research in Education* (pp. 103-50) Washington, D.C.: American Educational Research Association, 1983. [10]

Zumwalt, K. K. "Research on Teaching: Policy Implications for Teacher Education." In A. Lieberman and M. McLaughlin, eds., *Policy Making in Education: 81st Yearbook of the National Society for the Study of Education*. Chicago: University of Chicago Press, 1982. [50-51]